

医薬品 標的タンパク質 副作用の関連解析による
副作用の発現機序推定

水谷紗弥佳

目次

0.1	本稿での医薬品関連用語の使用	6
0.2	略語	7
第1章	全体への序論	8
1.1	薬物の副作用情報	8
1.2	薬物の標的分子	10
1.3	医薬品関連データベース	14
1.4	医薬品関連データ解析におけるバイオインフォマティクスの役割	16
1.5	本研究の目的と概要	17
1.5.1	本研究の目的	17
1.5.2	本研究の概要	17
第2章	研究Ⅰ 薬物 標的タンパク質相互作用ネットワークからの副作用の機序 推定と予測	20
2.1	序論	20
2.1.1	化学構造情報ベースのアプローチ	20
2.1.2	標的タンパク質情報ベースのアプローチ	21
2.1.3	システムワイドな解析法による副作用発現機序の推定と予測	22
2.1.4	本研究に至った経緯	25
2.2	方法	26
2.2.1	データセット	26
2.2.2	正準相関分析	27
2.2.3	スパース正準相関分析	29
2.2.4	予測精度の検証	32
2.2.5	エンリッチメント解析	32
2.3	結果と考察	34
2.3.1	薬物の標的タンパク質と副作用キーワードの相関成分の抽出	34
2.3.2	相関成分に抽出された標的タンパク質の集合の生物学的検証	36
2.3.3	抽出された相関成分の例	42
2.3.4	副作用予測への応用	45

2.3.4.1	予測精度の検証	45
2.3.4.2	副作用の予測	48
2.4	まとめと今後の展望	50
2.4.1	手法に関する展望	50
2.4.2	結果に関する展望	53
2.4.2.1	分子経路情報を用いたエンリッチメント解析の利点と限界	53
2.4.2.2	標的タンパク質の組織レベルでの発現の問題	56
2.4.2.3	薬物動態学的な副作用発現に関する展望	56
2.4.2.4	オフターゲットタンパク質への拡張	57
第3章	研究 II 薬物有害事象の疫学データを用いた副作用分類と薬剤疫学的特徴付け	58
3.1	序論	58
3.1.1	医薬品の自発報告システムの有用性と薬剤疫学	58
3.1.2	本研究に至った経緯	59
3.2	方法	61
3.2.1	市販後有害事象報告に登録された薬物と副作用のデータ	61
3.2.2	バイクラスタリング	64
3.2.3	Iterative Signature Algorithm (ISA)	64
3.2.4	バイクラスタの同定	65
3.2.5	副作用キーワードの医学的分類体系を用いた副作用クラスタの評価	69
3.2.6	薬物の適応症を用いた副作用クラスタの評価	69
3.2.7	患者の性別・年齢・体重の情報を用いた副作用クラスタの比較	69
3.3	結果と考察	70
3.3.1	バイクラスタリングによる副作用クラスタの同定	70
3.3.2	バイクラスタリング後の副作用と薬物の種類の評価	70
3.3.3	添付文書への参照による副作用クラスタの評価	72
3.3.4	副作用クラスタ内の副作用と薬物の特徴付け	73
3.3.5	副作用クラスタの例	75
3.3.6	患者の生理学的背景に基づいた副作用クラスタの特徴付け	81
3.3.6.1	第148クラスタと第163クラスタでの性別分布の比較	82
3.3.6.2	第106クラスタと第116クラスタでの年齢分布の比較	82
3.3.6.3	第108クラスタと第156クラスタでの体重分布の比較	83
3.4	まとめと今後の展望	88

3.4.1	方法に関する展望	88
3.4.2	結果に関する展望	89
第 4 章	全体のまとめ	91

表目次

1	1998 年以降に米国市場から撤退した市販薬 (Giacomini ら [22] より改変)	11
2	医薬品関連データベースの例	14
3	エンリッチメント解析の統計値 (Mizutani ら [45] より改変)	36
4	多くの相関成分 (CC) でエンリッチメントを示した KEGG パスウェイ マップ (Mizutani ら [45] より引用)	41
5	予測精度の比較 (Mizutani ら [45] より引用)	45
6	標的タンパク質プロファイルに基づき SCCA を用いて予測された副作用 (上位 20)	49
7	副作用の研究の比較 (Kuhn ら [37] より改変)	52
8	KEGG パスウェイマップ <i>Dopaminergic Synapse</i> に関する CC9 の標 的タンパク質	54
9	<i>Dopaminergic Synapse</i> 上の標的タンパク質に作用する CC9 の薬物	55
10	FAERS の 7 つのファイル構成	61
11	MedDRA の階層分類と用語数	62
12	副作用クラスタの例 (Mizutani ら [44] より改変)	79

目次

1	標的タンパク質の機能分類	13
2	標的タンパク質による薬物の分類	13
3	薬物に関連のある 3 つのデータ空間	23
4	スパース正準相関分析 (SCCA) による相関成分 (CC) の抽出	28
5	OCCA と SCCA の出力の比較	31
6	80 個の相関成分 (CC) に抽出された標的タンパク質と副作用キーワードのネットワーク表示 (Mizutani ら [45] より引用)	35
7	分子経路分類、あるいは、分子機能分類でのエンリッチメント解析	37
8	分子経路でエンリッチされた標的タンパク質の分子機能カテゴリ別分類 (Mizutani ら [45] より引用)	39
9	高い正準相関スコアを示した薬物の例 (Mizutani ら [45] より引用)	44
10	ROC 曲線を用いた予測精度の比較 (Mizutani ら [45] より引用)	46
11	Precision-Recall 曲線を用いた予測精度の比較 (Mizutani ら [45] より引用)	47
12	<i>Dopaminergic Synapse</i> を用いたドパミン伝達経路への薬理作用の模式図	54
13	Iterative Signature Algorithm (ISA) の手順	65
14	ISA アルゴリズムでのパラメータ最適化 1	67
15	ISA アルゴリズムでのパラメータ最適化 2	68
16	163 個の副作用クラスタの薬物と副作用の数 (Mizutani ら [44] より引用)	71
17	副作用クラスタにおける薬物の ATC 分類と副作用の MedDRA 分類 (Mizutani ら [44] より引用)	72
18	163 個の副作用クラスタ内の薬物と副作用の特徴付け (Mizutani ら [44] より改変)	74
19	副作用クラスタからの副作用の発現機序推定 (Mizutani ら [44] より引用)	80
20	副作用クラスタ間の比較	81
21	副作用発現の性別による違い (Mizutani ら [44] より改変)	85
22	副作用発現の年齢による違い (Mizutani ら [44] より改変)	86
23	副作用発現の体重による違い (Mizutani ら [44] より改変)	87

0.1 本稿での医薬品関連用語の使用

本稿で用いた用語は以下のように定義し、本文中で使い分けた。

薬物 (drug)

薬理作用を持つ化合物

薬剤 (agent)

有効成分としての薬物とその他の添加物などの混合物。経口、注射、吸入、外用などのために様々な投与のために異なる組成や性状を持つ。

医薬品 (medicine)

疾病の診断・治療・予防に使用される薬剤。

添付文書 (package insert)

医療用医薬品添付文書の略で、製品のパッケージに添付されている。

薬物療法 (medication)

薬物による疾病の治療。外科手術などと対比して用いる。

副作用 (side effect)

医薬品の使用に伴って生じる治療目的に沿わない作用全般。

有害反応 (adverse drug reaction)

副作用と同等の意味。

有害事象 (adverse drug event)

医薬品が適用された患者に生じたあらゆる好ましくない医療上の出来事で、必ずしも当該医薬品との因果関係があるわけではない。

適応症 (indication for use)

薬物の使用目的。治療目的。

0.2 略語

ADR (adverse drug reaction) : 薬物有害反応
ATC (Anatomical Therapeutic Chemical Classification System) : 解剖治療化学分類法
AUC (area under the ROC curve) : 受信者操作特性曲線下面積
AUPR (area under the PR curve) : 適合率 再現率曲線下面積
CC (canonical component) : 相関成分
CCA (canonical correlation analysis) : 正準相関分析
COX (cyclooxygenase) : シクロオキシゲナーゼ
CYP P450 (cytochrome P450) : シトクロム P450
EHR (electronic health record) : 電子健康記録・電子カルテ
FAERS (FDA Adverse Event Reporting System) : FDA 有害事象自発報告システム
FDA (US Food and Drug Administration) : 米国食品医薬品局
GABA (gamma-aminobutyric acid) : ガンマアミノ酪酸
GPCR (G protein-coupled receptor) : G タンパク質共役受容体
GPS (gamma Poisson shrinker) : シグナル検出法のひとつ
hERG (human *ether-a-go-go* related gene) : カリウムチャネル遺伝子
IC (information content) : 情報量
ISA (Iterative Signature Algorithm) : バイクラスタリングのアルゴリズム
MARTA (multi-acting receptor targeted antipsychotics) : 多受容体作用抗精神病薬
NSAID (non-steroidal anti-inflammatory drug) : 非ステロイド性抗炎症薬
OCCA (ordinary canonical correlation analysis) : 通常の正準相関分析
PR (precision-recall) : 適合率 再現率
PRR (proportional reporting ratio) : シグナル検出法のひとつ
ROC (receiver operating characteristic) : 受信者操作特性
RR (relative ratio) : シグナル検出法のひとつ
SCCA (sparse canonical correlation analysis) : スパース正準相関分析
SJS (Stevens-Johnson syndrome) : スティーブンス・ジョンソン症候群
SRS (spontaneous reporting system) : 自発報告システム
TEN (toxic epidermal necrolysis) : 中毒性表皮壊死症状
TNF (tumor necrosis factor) : 腫瘍壊死因子
VEGF (vascular endothelial growth factor) : 血管造血因子

第 1 章 全体への序論

医薬品としての薬物は、病気の症状を治療・緩和する効果（主作用）を持つが、一方で意図せぬ弊害（副作用）を引き起こすこともある。副作用には、眠気などの比較的軽いものから、呼吸困難や心血管イベント^{*1}などの重篤なものまである。また、多くの人に見られるものから、スティーブンス・ジョンソン症候群^{*2}のように少数の特定の人にもみ発現するものもある。薬物の副作用はいまだ高い罹患率、死亡率を記録している。Giacomini によると、米国では年間約 2 百万人の患者が重篤な副作用を発現しており、そのために死亡した患者の数は十万人にのぼる [22]。

副作用の発現は患者の生活の質を低下させるだけでなく、薬物治療の中断につながることもあり、医薬品の安全性を考える上で非常に重要な問題である。副作用には様々な発現パターンがあり、効率的な副作用の回避にはその作用機序を見極める必要がある。しかしながら、副作用発現の分子作用機序はほぼ未解明のままである。毒性や副作用を医薬品開発の段階で検出することは言うまでもなく重要であるが、規模や時間的な制約から見逃される副作用がある。そのため、開発段階で得られた知見や、市販後に調査された副作用報告の重要性が指摘されており、医薬品関連データベースが整備されてきている。これらのデータベースの情報を統合活用し、計算機科学的・統計学的解析手法を用いて、副作用の予測を行う研究が提案されてきている。しかしながら、副作用発現の機序の解明までには至っていない。本研究は、生体内分子のネットワーク構造を用いた解析を行い、副作用発現機序の解明と副作用予測を目指すものである。

1.1 薬物の副作用情報

医薬品の安全性は、開発段階においては、前臨床試験と臨床試験に分けて試験される。前臨床試験では培養組織や実験動物を用いた毒性試験を行い、次の臨床試験におけるヒトへの安全な投与可能性を検証する。臨床試験は第 I 相から第 IV 相で構成され、第 I 相から第 III 相までは「治験」と呼ばれる。治験ではまず第 I 相で薬物動態学的・薬力学的な視点から安全性の確認を行い、第 II 相で適切な用法・用量を決定した上、第 III 相では

^{*1} 心血管イベント (cardiovascular event) : 心筋梗塞や心不全など重篤な循環器系疾患の急激な発症や増悪を指す。

^{*2} スティーブンス・ジョンソン症候群 (Stevens-Johnson syndrome; SJS) : 皮膚や粘膜の過敏症で、死亡例もある。抗生物質や非ステロイド性抗炎症薬等、多くの医薬品が被疑薬とされている。

それまでに蓄積された安全性（と有効性）に関する知見を実際の患者の集団で検証する。前臨床試験と臨床試験で検出された有害反応や副作用は、医薬品の効能や用法とともに添付文書に記載され、医師や薬剤師への情報源となる。

承認を得て市場に出された後は第 IV 相が開始され、引き続き市販後調査が行われる。市販後調査は以下の理由で重要である。まず、治験においては対象患者数は数百人程度、用法・用量も画一的であり、投与期間も短い。そのため、まれな副作用は検出されない場合がある。また、治験では主に成人を対象とするため、小児、妊婦、高齢者での副作用は検出されにくい。一方、市販後調査では対象患者数も多く、用途や用量も様々であり、投与期間も長い。さらには、専門施設以外の医療機関でも広く使用されるため副作用の現れ方も多様である。従って、医薬品の安全性を適正に評価するためには、治験期間中に収集した情報だけでは不十分であり、市販後も継続して情報を収集することが重要となる。

実際、承認後に重篤な副作用が報告された事例が多数あり、表 1 に示すように、1998 年以降、19 個に及ぶ市販薬が重篤な副作用のために米国の市場から撤退した [22, 57]。このうち、副作用の発現機序が明らかになってきた薬物について例を 3 つ挙げる。

消化管機能改善薬であるシサプリド (Cisapride) と、抗ヒスタミン剤であるアステミゾール (Astemizole) は、Torsades de pointes^{*3} や QT 延長症候群^{*4} を引き起こすことが明らかになり、それぞれ 2000 年と 1999 年に市場から撤退した (表 1)。シサプリドはセロトニン受容体作動薬であり、アステミゾールはヒスタミン受容体拮抗薬であるが、共通のオフターゲット^{*5} として心臓のカリウムイオンチャネルである hERG への阻害作用を持つことが明らかにされた [18]。hERG 遺伝子の突然変異が QT 延長症候群などを引き起こすことは以前から知られていた。

ロフェコキシブ (Rofecoxib) は抗炎症剤・鎮痛剤であるが、長期使用により心筋梗塞などの心血管イベントの可能性が高まることが指摘され、2004 年に市場から回収された (表 1)。ロフェコキシブはシクロオキシゲナーゼ-2 (cyclooxygenase-2; COX-2) 選択的阻害剤であり、疼痛や炎症を引き起こすプロスタグランジン (prostaglandin) の産生を阻害する。その結果、プロスタグランジンの代謝物であり血液凝固抑制作用を持つプロスタサイクリン (prostacyclin) の産生も抑制する。Yu らはマウスを用いた実験で、プロスタサイクリン産生の抑制により凝血および血圧上昇が起こりやすくなることを明らかにし、心血

^{*3} Torsades de pointes : 心室頻拍で心室性不整脈の一種。

^{*4} QT 延長症候群 (long QT syndrome) : 心臓の収縮後の再分極の遅延がおき、Torsades de Pointes のリスクを増大させる心臓疾患。

^{*5} オフターゲット (off-target) : 開発時にデザインされた標的タンパク質の他に、薬物の作用が明らかになったタンパク質。

管イベントリスク増大との関連を示した [76]。

また、薬物を複数投与した場合にも副作用が発現・増強される場合がある。セリバスタチン (Cerivastatin) は高脂血症薬であるが、単独投与でも米国で死亡が出たばかりか、フィブラート系高脂血症薬ゲムフィブロジル (Gemfibrozil) との併用により横紋筋融解症*6 が多発したことが問題となり 2002 年に市場から回収された (表 1)。その後、ゲムフィブロジルとセリバスタチンとの間に薬物動態学的相互作用が起こることが明らかにされた。まず、2002 年に、ゲムフィブロジルがセリバスタチンの薬物動態に影響を及ぼすことが発表された [5]。肝ミクロソームを用いた研究では、ゲムフィブロジルによるセリバスタチンのグルクロン酸抱合化抑制により、セリバスタチンの血中濃度が増加し副作用を引き起こすことが示唆された [54]。さらに、2004 年には、ゲムフィブロジルとそのグルクロン酸抱合体がシトクロム P2C8 (cytochrome P2C8; CYP2C8) を強く阻害するとともに、セリバスタチンの肝臓への取り込みに関与する輸送体である有機アニオン輸送ポリペプチド (organic anion transporting polypeptide 2; OATP2) をも阻害することが明らかにされた [58]。

1.2 薬物の標的分子

薬物は体内の作用点に輸送され、標的とする分子に作用する。標的分子への薬理作用の種類により、作動薬、拮抗薬、遮断薬、阻害薬などに分類される。薬物の標的分子は主にタンパク質である。他にも DNA、RNA、生体内化合物、糖鎖分子などがあるが、本研究では触れない。ヒトゲノムプロジェクトによりゲノムにコードされた遺伝子の注釈が網羅的に付けられるようになり、ゲノム情報からの標的タンパク質の同定が急速に進んだ。そこで、「標的分子の数はどれくらいあるのか」という議論が起こり、Overington らは承認薬に対してその標的分子の数を 324 個と見積もっている [50]。しかしながら、集計の仕方は研究者やデータベースによってまちまちである。

KEGG DRUG [33] には 2014 年現在、薬物の標的分子として 803 個のヒトタンパク質が登録されている。分子機能分類システムである KEGG BRITe [33] を用いてこれらを 8 つのタンパク質カテゴリに分類すると図 1 のような分布になる。全体の 33% が酵素 (Enzymes) であり、17% が G タンパク質共役型受容体 (G protein-coupled receptor; GPCR)、12% がイオンチャネル (Ion channels) に分類され、他にタンパク質キナーゼ (Protein kinases)、サイトカイン (Cytokines) とサイトカイン受容体 (Cytokine

*6 横紋筋融解症 (rhabdomyolysis) : 筋力低下や筋肉痛を伴う骨格筋細胞の壊死。

表 1 1998 年以降に米国市場から撤退した市販薬 (Giacomini ら [22] より改変)

医薬品名	承認	撤退	適応症	危険性
Mibefradil	1997	1998	高血圧 狭心症	薬物間相互作用 Torsades de pointes
Bromfenac	1997	1998	非ステロイド性抗炎症薬	急性肝不全
Terfenadine	1985	1998	抗ヒスタミン剤	Torsades de pointes 薬物間相互作用
Astemizole	1988	1999	抗ヒスタミン剤	Torsades de pointes 薬物間相互作用
Grapfloxacin	1997	1999	抗生物質	Torsades de pointes
Etretinate	1986	1999	乾癬	先天的欠損症
Alosetron	2000	2000 (2002) [†]	女性の過敏性腸症候群	虚血性大腸炎 慢性便秘
Cisapride	1993	2000	胸やけ	Torsades de pointes 薬物間相互作用
Troglitazone	1997	2000	糖尿病	急性肝不全
Cerivastatin	1997	2001 (2002) [†]	高脂血症	横紋筋融解症 薬物間相互作用
Rapacuronium	1999	2001	麻酔薬	気管支痙攣
Levomethadyl	1993	2003	アヘン依存	致命的な不整脈
Rofecoxib	1999	2004	鎮痛剤	心臓発作 脳卒中
Valdecoxib	2001	2005	鎮痛剤	皮膚反応 (SJS)
Natalizumab	2004	2005 (2006) [†]	多発性硬化症	脳感染
Technetium (99m Tc) fanolesomab	2004	2005	診断補助薬	心肺停止
Pemoline	1975	2005	注意欠陥多動性障害	肝不全
Pergolide	1988	2007	パーキンソン病	弁膜症
Tegaserod	2002	2007	便秘を伴う 過敏性腸症候群	狭心症 心臓発作 脳卒中

[†] 制限付きで再販売

receptors)、輸送体 (Transporters)、核内受容体 (Nuclear receptors) と続く。図 2 は、これらを標的とする薬物の数の分布図である (図 2)。やはり酵素 (Enzymes)、GPCR、イオンチャネル (Ion channels) に作用する薬物が大半を占める。

酵素のカテゴリと GPCR カテゴリから代表的な例を挙げる。酵素の阻害薬として代表的な薬物に非ステロイド性抗炎症薬 (non-steroidal anti-inflammatory drugs; NSAIDs) がある。NSAID はシクロオキシゲナーゼ阻害薬で、痛みや炎症を惹起するプロスタグランジンの産生を阻害することによる鎮静効果を持つ。GPCR に作用する薬物の例としては、ドパミン受容体拮抗薬などの抗精神病薬が挙げられる。ドパミン受容体 D2 は G タンパク質と共役し、ドパミン結合により cAMP 産生を抑制する。統合失調症ではその結果として、攻撃性などの陽性症状が現れると考えられているが、抗精神病薬は内在性ドパミンと競合することでこの攻撃性を押さえる効果がある。

薬物の副作用の機序を研究する上で、標的タンパク質への作用は非常に有用な情報である。本研究に関連しては、以下の 2 点が特に重要であるため、ここで言及しておく。

1 点目は、薬物の多くは標的タンパク質に対してそれほど高い選択性を持たないばかりか、非特異性 (promiscuity)^{*7}を表すものが多数あると言われていることである [50]。そのため、前述のカリウムイオンチャネル hERG のように、オフターゲットへの作用が副作用につながるものが指摘されている。

2 点目は、薬物の標的タンパク質に限らず、タンパク質は細胞内外で互いに情報のやりとりをする生化学経路を形成し、生体機能を複雑に制御しているため、一種類のタンパク質への制御の影響が広い範囲に渡って波及することである。前述の COX-2 阻害剤の例のように、標的タンパク質への制御は、その直接の生産物の産生抑制ばかりでなく下流の代謝物の産生をも抑制し、その結果、本来は意図せぬ心血管イベントのような副作用を引き起こすことがある。

^{*7} 非特異性 (promiscuity) : 薬物が複数の標的分子に非特異的に作用すること。そのような薬理作用を多重薬理 (polypharmacology) という。

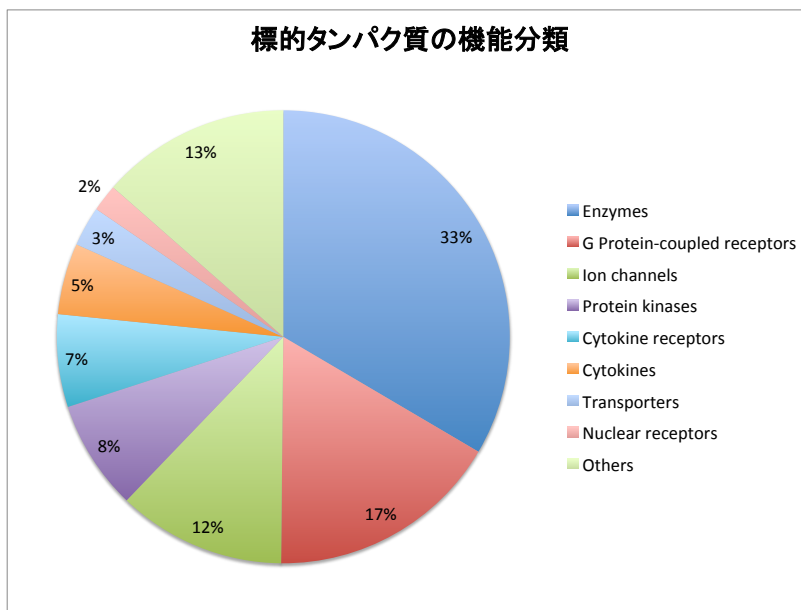


図1 標的タンパク質の機能分類

KEGG DRUG に登録された薬物の標的である 803 個のヒトタンパク質を KEGG BRITE の分子機能分類システムを用いて 8 つのカテゴリに分類した (n = 803)。ただし Enzymes と Protein kinases など一部重複がある。

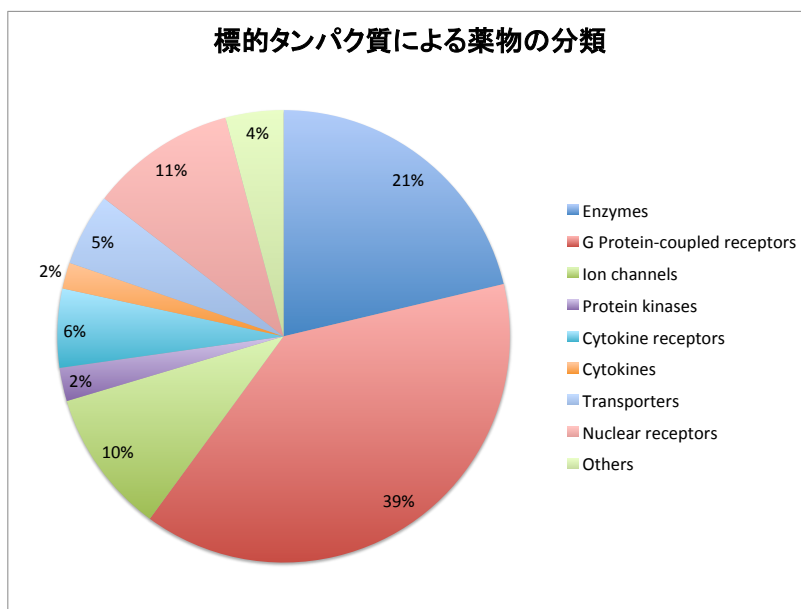


図2 標的タンパク質による薬物の分類

KEGG DRUG に登録された薬物を標的タンパク質の分子機能カテゴリで分類した (n = 3392)。

1.3 医薬品関連データベース

医薬品に関する様々な知識や情報が公開文書や医療用医薬品添付文書（以下、添付文書と略す）あるいは学术论文から抽出され、公共のデータベースに蓄積されている。医薬品関連データベースはそれぞれ異なる目的やポリシーに基づいて作られ、下記の4つのタイプに大別される（表2）。

表2 医薬品関連データベースの例

データベース	本研究で使用	URL	参考文献
(1) 総合情報データベース			
DrugBank		http://www.drugbank.ca/	[36]
KEGG DRUG		http://www.genome.jp/kegg/	[33]
DailyMed		http://dailymed.nlm.nih.gov/	該当無し
PharmGKB		http://www.pharmgkb.org/	[67]
(2) 化合物データベース			
PubChem		http://pubchem.ncbi.nlm.nih.gov/	[41]
(3) 標的タンパク質データベース			
Matador		http://matador.embl.de/	[26]
ChEMBL		https://www.ebi.ac.uk/chembl/	[10]
SuperTarget		http://bioinf-apache.charite.de/supertarget.v2/	[28]
TTD		http://bidd.nus.edu.sg/group/cjttd/	[55]
(4) 副作用データベース			
SIDER		http://sideeffects.embl.de/	[38]
FAERS		http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/	該当無し

(1) 総合情報データベース

このタイプのデータベースには、DrugBank[36]、KEGG DRUG、DailyMed (<http://dailymed.nlm.nih.gov/>)、PharmGKB[67] が含まれ、承認薬の添付文書に記載された薬物の化学的性質、薬効および用法、標的分子などを包括的に登録してい

る。DrugBank は、薬物の化学情報、薬理学情報、商品情報など 200 以上のデータ属性に渡って情報を蓄積した包括的な医薬品データベースである。特にタンパク質に関するデータを豊富に記載し、7,740 個の薬物エントリに関する 4,282 個のタンパク質（標的分子/酵素/トランスポーター/担体タンパク質）の情報を提供している。KEGG DRUG は、日本、米国、欧州で承認された薬物の化学成分、化学構造、標的分子、薬物代謝酵素、および、薬物に関する分子間相互作用ネットワーク情報を包括的に蓄積したデータリソースであり、10,096 個の薬物エントリを登録している。

(2) 化合物データベース

このタイプのデータベースには PubChem [41] などがあり、薬物を含む化合物の化学構造や化学的性質に関する情報を収集提供している。

(3) 標的タンパク質データベース

このタイプのデータベースには ChEMBL [10], Matador [26], SuperTarget [28], TTD [55] などがあり、標的分子への結合、活性制御、薬理効果などを収集提供している。

(4) 副作用データベース

このタイプのデータベースには、添付文書に記載された副作用情報を収集したデータベース (SIDER [38] など) と、市販後調査から報告された有害事象を収集したデータベース (FAERS (<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/>) など) がある。SIDER は、医薬品の公開文書と添付文書から抽出された副作用情報を蓄積したデータベースで、996 個の薬物に関して 4,192 個の副作用キーワードとその発現頻度を対応付けている。FAERS は、米国食品医薬品局 (US Food and Drug Administration; FDA) が提供している有害事象の自発報告システム (spontaneous reporting system; SRS) であり、薬物による有害反応 (adverse drug reaction; ADR) の報告を大規模に収集している。2014 年現在 5 百万件を超える報告を公開している。ADR と被疑薬だけでなく、薬物の適応症 (indication for use) (薬物治療の目的となる患者の病気や症状) などの付随データ、および、患者の性別・年齢・体重などの生理学データなどを含み、薬剤疫学的に非常に有用なデータベースである。

各々のデータベースは、計算処理を可能にする薬物へのインデックスや、付随データに関する詳細な情報 (標的タンパク質のアミノ酸配列など) あるいは外部データベースへ

の参照情報などが付加されており、統合的に利用することにより利用価値が高まる。一方、薬物名の記法が統一されていない点など、利用上注意を要する点も挙げられる。本研究ではそれぞれのデータベースの特性と相互参照の容易さを考慮して、上記のデータベースのうち、DrugBank、KEGG DRUG、Matador、PubChem、SIDER、FAERS を利用した。

1.4 医薬品関連データ解析におけるバイオインフォマティクスの役割

医薬品関連データベースには、薬物の化学構造、標的分子、薬物代謝酵素、相互作用薬物、薬効（主作用）、副作用など様々な属性のデータが登録されている。これらのデータは、薬物投与に関する情報を異なる視点で収集したデータと見なすことができる。例えば、標的分子の多くは生体が内在的に持つタンパク質で、その機能は分子経路などを通じて複雑に調節されている。薬物は、このような生体内物質に作用し、その機能調節に介在する外因性の（化学）物質と見なすことができ、効能や副作用はその結果として現れる生体の表現型と見なすことができる。このような視点で見ると、医薬品関連データは、生体情報を異なる解像度で記述した事実の集積であり、データ間の重要な「関係」を見いだすことは、薬物が引き起こす表現型の機序を分子レベルで推定することにつながる。

バイオインフォマティクスの手法は、様々なデータ属性を持つ大規模なデータの整理と標準化、大規模データからの未知の事例の発見（マイニング）と、発見された事例の統計的有意性の検定、既存の知識による注釈と分類、可視化などに役立つ。特に、機械学習の手法を取り入れたデータ間の関連解析は、性質の異なるデータ間でも顕著に観察される「関係性」を検出するのに非常に有効である。

Yamanishi らは、代謝経路の再構築を目的として、遺伝子発現データ、遺伝子位置データ、遺伝子系統プロファイルデータなどのデータ構造の異なる複数のデータを機械学習の方法を用いて遺伝子ネットワークとして統合し、遺伝子間の機能的関連性を予測する方法を提案している [75]。さらに、Yamanishi らはこの方法を医薬品関連データの解析にも応用し、薬物の化学構造の類似性と標的タンパク質のアミノ酸配列の類似性から、薬物とタンパク質の潜在的な相互作用を予測する研究を行った [73]。このように、医薬品関連データへの関連解析の有効性を示した先行研究がいくつかある [4, 8, 12, 19]。

1.5 本研究の目的と概要

1.5.1 本研究の目的

今までの研究では、化学構造や、標的分子への直接的な薬理作用のデータを中心に薬物の特徴付けがなされてきた。しかしながら、標的分子への薬理作用の影響は、生化学経路、細胞、組織、臓器など複雑な階層構造を通じて生体全体へ波及するため、薬物が標的分子を共通に持たない場合でも間接的に（同じ代謝経路に作用するなどして）類似した副作用が発現される場合がある [6]。また、類似した副作用を引き起こす薬物でもその発現に至る生化学経路は異なる場合もある。従って、薬物と副作用の関係をその発現機序も含めて包括的に解析し理解することは、副作用対策にも関わる重要な研究課題である。近年、添付文書などに記載された薬理的、毒性学的な知識を蓄積した医薬品関連データベースのデータから、バイオインフォマティクス的手法を用いた解析により、副作用発現の予測が試みられてきた。しかし、発現機序の推定まで踏み込んだ研究は少ない。本研究では、副作用発現の機序推定と未知の副作用発見の双方を目的として新しいデータ解析法を提案する。

研究 I（第 2 章）では、薬物と標的タンパク質の相互作用情報を用いることが副作用発現の機序推定を可能にするのではないかとこの立場から、薬物と標的タンパク質の相互作用情報に基づき副作用を予測する解析法を提案した [45]。

研究 II（第 3 章）では、副作用発現機序の推定には、治験データだけでなく、個人差をより反映した大規模な疫学データが重要であるとの立場から、医薬品の市販後報告データを用いて患者の特性と副作用との関連を調査した [44]。

1.5.2 本研究の概要

研究 I の概要

背景と目的：既存の研究では、薬物の化学構造情報に基づく副作用予測のアプローチがとられてきた。この背景には、類似した化学構造を持つ薬物は類似した分子に結合し、オフターゲットへ薬理作用を持つことが副作用の原因であるという考え方がある。しかしながら、生体内の標的分子は複雑な分子ネットワークにより制御されているため、化学構造の異なる薬物でも、同一の分子経路を介して間接的に副作用を共有するようなケースもあり、このようなケースは化学構造情報ベースのアプローチでは検出が難しい。研究 I では、薬物の標的タンパク質との相互作用情報を用いることによって、間接的な副作用発現

も検出できる解析法を開発し、分子ネットワークの視点から副作用発現の機序に関する新たな知見の獲得を目的とした。特に、異なる化学構造を持つ薬物が副作用を共有するケースについて特に注意して解析することを目的とした。また、提案した解析法を応用して、標的タンパク質情報に基づき薬物の副作用を予測する方法を提案した。

方法：市販薬の添付文書に記載された副作用情報を収集した SIDER データベースから薬物と副作用キーワードとの対応表を取得し、標的タンパク質情報を登録した DrugBank データベースと Matador データベースから薬物と標的タンパク質との対応表を取得した。次に、機械学習のひとつであるスパース正準相関分析 (Sparse Canonical Correlation Analysis; SCCA) により、標的タンパク質と副作用という特徴量が有意なレベルで共通点を持つ薬物のグループを同定した。このとき、薬物 標的タンパク質 副作用キーワードという 3 つの要素の集合を持つグループが複数同定され、各グループの標的タンパク質と副作用キーワードは有意に相関する。このグループを相関成分 (canonical component; CC) と呼ぶことにする。同定された相関成分の評価は、KEGG データベースと Gene Ontology データベースから標的タンパク質の分子経路分類と分子機能分類の情報を得てエンリッチメント解析を行った。また、標的タンパク質情報に基づいた副作用予測法の精度を既存の化学構造情報に基づいた予測法と比較した。

結果：80 個の CC を同定し、各 CC に抽出された標的タンパク質の集合について、エンリッチメント解析を行った。その結果、それらの標的タンパク質群は、分子機能分類上よりも分子経路分類上でのまとまりがより強い傾向にあることが確認された。このことから、副作用の発現機序が分子経路でより顕著に特徴付けられることが示唆された。標的タンパク質プロファイルに基づいた予測法を副作用未知の薬物に適用した結果、化学構造情報を用いた既存のアプローチより良い予測精度を示した。また、高いスコアで予測された副作用のいくつかについては文献調査により妥当性を確認することができた。

研究 II の概要

背景と目的：添付文書から得られる情報は、前臨床での毒性検査および臨床試験での小規模な患者集団での発現が報告された副作用に限られるため、個人差による発現を反映していない。患者の多様性を加味した副作用発現の理解には、添付文書の記載だけでは不十分であり、より大規模な疫学データを解析する必要がある。既存の研究でも、このような疫学データから、薬物と副作用の組み合わせのうち顕著に報告されるものを検出するシグナ

ル検出法が開発されてきたが、必ずしも副作用発現の機序には着目していない。さらに、個人差による発現パターンの検出はほとんど手つかずのままである。研究 II では、このような発現の傾向を持つ副作用も含めた方法の開発を目的とした。

方法：FAERS の大規模な薬物有害事象報告データから、副作用を行とし薬物を列とする報告頻度行列を作成した。データマイニングの方法のひとつであるバイクラスタリングを用いて、発現頻度の類似した副作用と、それに関係する薬物のクラスタを同定した。このクラスタを副作用クラスタと呼ぶことにする。医薬用語の分類体系である MedDRA の階層分類と、薬物の適応症データを用いて、同定された副作用クラスタを評価した。副作用クラスタ内の薬物と副作用の作用機序を KEGG DRUG、DrugBank、DailyMed などのデータベースを参照して推定した。FAERS に登録された患者の性別・年齢・体重など生理学的背景の情報を用いて、同定された副作用クラスタを特徴付けた。

結果：バイクラスタリングの結果、163 個の副作用クラスタを同定した。各副作用クラスタ内に現れる副作用の集合を MedDRA の階層分類を用いて評価した結果、ほとんどの副作用クラスタで副作用の集合の医学的類似性が確認された。各副作用クラスタ内に現れる薬物の集合を、その適応症（効能）情報を用いて評価したところ、163 個の副作用クラスタのうち 145 個（89%）では、副作用が薬物の適応症に非依存的に発生している可能性が示唆された。この 145 個の副作用クラスタのうち代表的な 3 つについて、その作用機序を推定した。さらに、患者の生理学的情報を用いてクラスタを特徴付けたところ、性別・年齢・体重の分布に顕著な違いがあった。

第2章 研究Ⅰ 薬物 標的タンパク質相互作用ネットワークからの副作用の機序推定と予測

2.1 序論

副作用に関係するタンパク質は数百に及ぶと言われている [42]。前臨床試験では、副作用との関係のよく知られているタンパク質に対して新規化合物の親和性を検証するアッセイ系が確立されているが、費用の問題から *in silico* での（計算機による）副作用予測手法が開発が進められてきた。その主な目的は、創薬の早い段階で新規化合物の毒性や潜在的な副作用を予測するという非常に実践的なものである。しかしながら、この方法では副作用の発現機序を解明することはできない。

薬物は標的分子の活性を制御するにとどまらず、細胞、組織、臓器あるいは個体レベルで多様な作用を持つため、異なるレベルでの実験を重ね、副作用発現を網羅的に解明することは非常に時間と労力がかかる。したがって、薬物に関する情報を生体内分子に関する情報とともにデータベースに蓄積、統合利用することにより、副作用に関する新知見を発見する手段として、計算機を用いたシステムワイドな解析法の開発が求められている。そのようなアプローチとして関連解析が挙げられる。関連解析では、医薬品に関する異なる属性のデータ（化学構造、標的タンパク質、副作用など）ごとに既知の情報を収集し、これらのデータ間で顕著に現れる「関係」を抽出する手法である。特に、生体の表現型レベルのデータである副作用データと、分子レベルのデータである標的タンパク質のネットワークデータの間の重要な関係を抽出することで、副作用発現の俯瞰的な理解が可能になる [19, 20, 61]。副作用の予測に関する既存のアプローチを以下に挙げる。

2.1.1 化学構造情報ベースのアプローチ

化学構造情報ベースのアプローチは、薬物の化学構造と標的分子への作用との相関関係を明らかにし、未知の標的分子への作用の予測を目的としている。薬物の多くは分子量の小さい化学物質であるため、原子や修飾基、分子結合の有無などを記述するフィンガープリントと呼ばれる記述子を用い、薬物の化学構造の特徴を記述することが可能である。これを学習データとして統計モデルを学習させ、標的分子を予測する方法が提案されてきた [35, 47, 74]。

化学構造情報ベースのアプローチは副作用の解析にも応用されている。この応用は、類似した化学構造を持つ薬物は類似した化学的性質や薬理的性質を持つため、類似した副

作用を引き起こすのではないかという考え方に基づいている。実際、ノバルティスファーマ社の Scheiber らにより、米国での承認薬に関して、化学構造に基づく副作用の類似性が検証された [57]。

その後 Atias と Sharan により、正準相関分析 (CCA) を用いて化学構造情報から副作用を予測する予測法が提案された [4]。また Pauwels らは、従来の正準相関分析を改良したスパース正準相関分析 (SCCA) を用いて化学構造情報から副作用を予測する方法を提案した [52]。ここに挙げたアプローチは、化学構造と副作用との直接的な対応のみを情報として副作用予測を行っている。

類似した化学構造を持つ薬物が類似した副作用を引き起こすという考え方の背景には、共通のオフターゲットへの作用が仮定としてがある。Campillos らは、共通のオフターゲットの検出を目的として、市販薬の添付文書に記載された副作用情報を元に、副作用プロファイルの類似した薬物ペアを予測した。その結果として得られた薬物ペアの約 2 割は化学構造の類似性の低い薬物であり、オフターゲットの共有は化学構造の類似性に必ずしも限定されないことを示唆した [12]。

実際、化学構造の類似した薬物が必ずしも同程度の毒性を持つとは限らない。その例として脂質異常症治療薬であるスタチンの開発の歴史を挙げると、1973 年にアオカビから発見された最初のスタチンであるメバスタチンは重篤な副作用の発生のため承認には至らなかったが、後にコウジカビから発見されたロバスタチンと、合成薬であるプラバスタチンは比較的高い安全性を持っていたため商品化に成功した。これらの承認約とメバスタチンとの構造的な違いは、メチル基 (ロバスタチン) あるいは水酸基 (プラバスタチン) のみである。

2.1.2 標的タンパク質情報ベースのアプローチ

次に、薬物の標的タンパク質に注目して副作用との関連を研究するアプローチを紹介する。タンパク質情報ベースのアプローチの利点は、タンパク質に関する生物学的な情報を豊富に用いることができる点にある。実際、分子生物学、遺伝学、生化学、発生生物学、タンパク質工学など様々な研究分野により得られた膨大な知見がデータベースに蓄積され、計算機による解析が可能である。

前述のように、副作用発現に関与するタンパク質は数百に及ぶと言われ [42]、その機序には次のような知見が定着している。まず、主要な標的分子の制御 (modulation) による発現である。特定の副作用と強い関連があることが知られている標的タンパク質の例として、1.1 で紹介した COX-2 阻害剤による心血管イベントのリスク [76] や、セロトニン受容体 5-HT_{2B} (5-Hydroxytryptamine receptor 2B) による心弁膜症 (cardiac

valvulopathy) のリスク [68] が挙げられる。また、プレグナン X レセプター (pregnane X receptor; PXR) は、リガンド結合領域の可動性が高いため、様々な薬物作用による副作用の可能性が指摘されている [17]。PXR は薬物代謝酵素シトクロム P450 の一種である CYP3A4 の発現を誘導する核内受容体であるため、薬物代謝に関係する副作用の可能性が考えられる。次に、医薬品の開発段階では認識されず、市販後に発生した副作用によりオフターゲットとして発覚した例がある。1.1 でも紹介したように、カリウムイオンチャンネル hERG は良く知られたオフターゲットである。

近年、*in silico* の方法によりオフターゲットの探索を目的とした研究が盛んに行われている。2.1.1 でも述べたように、Campillos らは、市販薬の添付文書に記載された副作用情報を元に副作用プロファイルの類似した薬物ペアを同定し、その結果としてオフターゲットを検出した。オフターゲット探索の代表的なアプローチとして、リガンドのタンパク質結合情報を用いるアプローチがある。Lounkine らは 656 個の薬物に対し、Similarity Ensemble Approach (SEA) [34] と呼ばれるターゲット予測手法を用いて、副作用との関連の知られている 73 個のタンパク質の中からオフターゲットとなるものを探索し、潜在的な副作用との関連付けを行った [42]。SEA は薬物の化学構造情報を元にタンパク質への結合可能性を予測するが、Lounkine らの研究ではその補足的なデータとして、ChEMBL データベースに登録された結合親和性の情報を用いている。

オフターゲット探索のもうひとつの代表的なアプローチは、タンパク質立体構造情報を用いたアプローチである。Xie らは、ドッキングの手法を用いて、薬物のオンターゲットタンパク質と同様の結合ポケットを持つオフターゲットタンパク質を同定し、既知の生物学的経路にマッピングすることで潜在的な副作用を予測した [71]。また、Wallach らも同様のドッキングアプローチを使用しオフターゲットを探索し、さらにその発展として、タンパク質の分子経路情報を用いて、分子経路の視点から副作用の発現機序の推定を行った [66]。しかしながら、ドッキングベースの方法はタンパク質立体構造の利用可能性に大きく依存するため、立体構造情報の少ない膜タンパク質などへの応用は制限される。

2.1.3 システムワイドな解析法による副作用発現機序の推定と予測

最後に、副作用の発現機序の推定と副作用予測を、よりシステムワイドに行った研究を紹介する。前述の 2 つのアプローチからの発展は、薬物、標的タンパク質、副作用という 3 つの属性のデータをそれぞれ、化合物空間 (chemical space)、生物空間 (biological space)、副作用空間 (side effect space) という集合とみなし、関連解析を用いて 3 つの空間の間の相関関係を検証するという概念の構築にある (図 3)。この見方をすることで、薬物 標的分子 副作用という一連の因果関係を推定することができ、薬物の副作用発現

の機序解明へさらに一步近づくことができる。

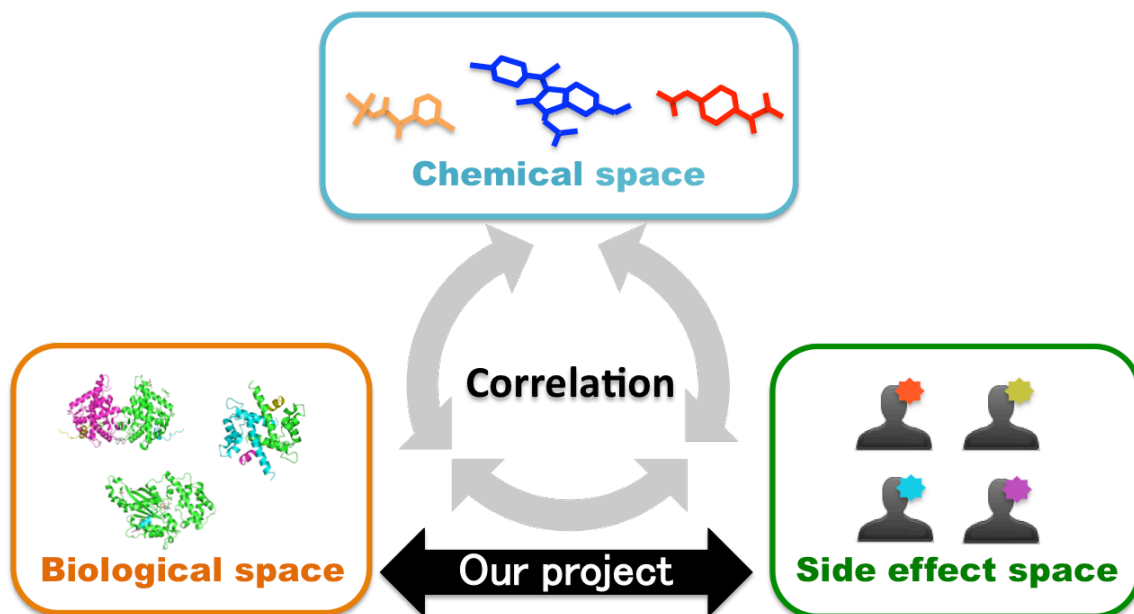


図3 薬物に関連のある3つのデータ空間
薬物の集合としての化合物空間 (chemical space)、標的タンパク質の集合としての生物空間 (biological space)、副作用の集合としての副作用空間 (side effect space)

以下に挙げる研究では、上の3つのデータ属性のうち、2つの属性からサンプルを取り、両者の間の相関を検証している。

ノバルティスファーマ社の Bender らは、70 個の標的タンパク質と 166 の副作用について、各組み合わせ (標的タンパク質 副作用のペア) の類似性を測るために、4,890 化合物の化学構造 (フィンガープリント) を用いてマルチカテゴリベイズモデルを学習させ、有意に相関のある組み合わせを予測した [8]。ケーススタディとして、 μ オピオイド受容体、ムスカリン M2 受容体、および、COX-1 と関連の強い副作用について論じている。しかしながら、標的タンパク質を介した副作用発現機序の網羅的な推定を目的とする解析は行われていなかった。

その後 Scheiber らは、同じくマルチカテゴリベイズモデルを用いて予測された標的分子 副作用間の相互作用をパスウェイのデータベースである MetaBase (GeneGo) にマップすることで、分子経路レベルでの副作用発現機序推定を行っている [56]。

さらに、分子レベルでの「薬物と標的タンパク質の相互作用」と、表現型レベルでの「薬物と副作用の相関」という2つの異なるレベルでの薬物作用データを統合することで

薬の副作用を予測するためのシステムワイドな解析法が取られるようになってきた（図3）[19, 20, 61]。

ファイザー製薬の Fliri らのグループは、1,042 個の薬物に対して非臨床と臨床の視点から薬物の作用を比較した [19]。非臨床の視点からは、生物学的スペクトル (biological spectra) と呼ばれる 92 個のタンパク質に対しての阻害実験によるプロファイルを作成し、臨床の視点からは、副作用スペクトル (side effect spectra) と呼ばれる副作用のプロファイルを作成して、それぞれのデータ (スペクトル) に階層的クラスタリングを実行したところ、40 個の薬物において類似した副作用プロファイルを表す薬のグループが同じような結合性を持つ事が認められた。特に興味深い知見として、ベンゾジアゼピン (benzodiazepine) 系の薬物とゾピクロン (Zopiclone) は化学構造は似ていないが、どちらも GABA_A 受容体への結合性を示した。この2種類の薬物は副作用の類似性も示しており、従来の化学構造の類似性による副作用予測では得られないような知見が得られた。

さらに、薬の作用を予測する方法として、ここで提案された標的タンパク質ベースの方法と化学構造情報のプロファイルベースの方法とを比較した結果、化学構造ベースのプロファイルと副作用との相関関係も見られるものの、データサイズや比較方法によらず標的タンパク質ベースの方法の方が良い性能を示すことが分かった [20]。

2.1.4 本研究に至った経緯

薬物による内在性の分子への機能制御は、生化学経路、細胞、組織、臓器などの階層化された機構を通じて生体内の広い範囲に影響を及ぼすため、生体内をシステムワイドに解析する研究手法の開発が求められている [6]。しかしながら、大規模データセットで薬物の標的タンパク質と副作用を実験的に決定することは費用の面でも効率の面でも未だに難しい [68]。従って、計算機科学的な方法の開発が求められており、その出発点として、上記に挙げたような生化学経路を視野にいれた研究が発表されている。その結果、副作用の発現には、オフターゲットへの直接の薬理作用だけでなく、標的タンパク質を含めた分子経路全体への作用が寄与しているという見解が一般的になりつつあるが、その実証はまだない。

これらのことを背景として、本研究では、2.1.3 の概念に則り、図 3 に示した 3 つのデータ空間の間の相関関係を検証した。特に、標的タンパク質と副作用の關係に注目し、「薬物とタンパク質の相互作用」という分子レベルでの相互作用と、「薬物と副作用の対応」という表現型レベルでの相互作用という解像度の異なる相互作用データを統合することで、「薬物 標的タンパク質 副作用」という一連の機序を推定することを目的とした。本研究では、「薬物 標的タンパク質 副作用」という 3 つのデータ空間の間で重要な関係にある多対多の組み合わせを、機械学習の手法のひとつであるスパース正準相関分析を用いて相関成分として同定した。同定された相関成分を分子経路や分子機能の情報をを用いて評価した。特に、異なる化学構造を持つ薬物が共通の副作用を発現する場合の機序を明らかにすることを目的として、副作用を共通に持つ標的タンパク質の集合を分子経路の観点から特徴付けた。さらに本解析手法を応用して、標的タンパク質プロファイルから副作用を計算機科学的に予測する方法を提案した。

2.2 方法

2.2.1 データセット

薬物の標的タンパク質プロファイル、副作用プロファイル、化学構造プロファイルを以下の手順で作成した。

DrugBank データベースと Matador データベースから薬物の標的タンパク質情報を取得した。これらのデータベースは薬理効果を持つことが確認されている標的タンパク質のみならず、薬物と直接あるいは間接的に相互作用することが知られているタンパク質も含んでいる。データベースには薬物代謝酵素であるシトクロム P450 や不特定のタンパク質に結合するアルブミンなども含まれていたが、これらを標的タンパク質と明確に区別することは困難であったため、標的タンパク質への薬力学的作用に焦点をおく本研究では、これらのタンパク質はデータセットから除いた。副作用情報は、市販薬の添付文書に記載された副作用キーワードを蓄積した SIDER データベースから取得した。

合計で 658 個の薬物が標的タンパク質情報と副作用情報を共に持ち、1,368 個の標的タンパク質 (5,074 対の薬物と標的タンパク質への作用) および、1,339 個の副作用キーワード (49,051 対の薬物と副作用キーワードの対応) を含んでいた。この 658 個の薬物の各々について、標的タンパク質との相互作用の有無を 1 または 0 で表した 1,368 次元標的タンパク質プロファイル、および、副作用キーワードとの対応有無を 1 または 0 で表した 1,339 次元副作用プロファイルを作成した。

Atias と Sharan の研究 [4]、および、Pauwels ら研究 [52] により提案された化学構造情報ベースの副作用予測と比較するために、本研究で解析対象とする 658 個の薬物の化学構造プロファイルも作成した。PubChem データベースで定義された 881 種類の化学部分構造から構成されたフィンガープリントを用いて、658 個の各薬物に部分構造の有無を 1 または 0 で表した 881 次元の化学構造プロファイルを作成した。

2.2.2 正準相関分析

同一のサンプル（ここでは薬物）に関して 2 種類のデータセット（標的タンパク質プロファイルと副作用プロファイル）が与えられたとき、この 2 種類のデータセットの間の相関関係を解析する統計手法として正準相関分析（canonical correlation analysis; CCA）が考えられる。正準相関分析では、標的タンパク質と副作用キーワードを特徴と呼ぶ。正準相関分析のアイディアは、2 種類の特徴の集合の中から、多くのサンプルに共起する特徴を出力することである。

正準相関分析はもともと、1936 年に Hotelling により提唱された多変量解析の手法であるが、近年、機械学習の分野で改良が進み、スパース正準相関分析（sparse canonical correlation analysis; SCCA）が開発された [70]。

SCCA を用いて、高い相関関係にある標的タンパク質と副作用キーワードの集合を抽出するためのワークフローを図 4 に示す。まず、薬物と標的タンパク質の相互作用（図 4A の “Molecular interaction”）の有無をそれぞれ 1 と 0 で表す二値行列を作成する（図 4B）。薬物と副作用キーワードの対応（図 4A の “Phenotypic interaction”）に対しても同様の二値行列を作成する。

次に、この 2 つの行列に SCCA を適用して、顕著に多くの薬物に共起している標的タンパク質と副作用キーワードの集合を抽出する。この集合を相関成分（canonical component; CC）と呼ぶ（図 4C）。相関成分は複数存在しうる。

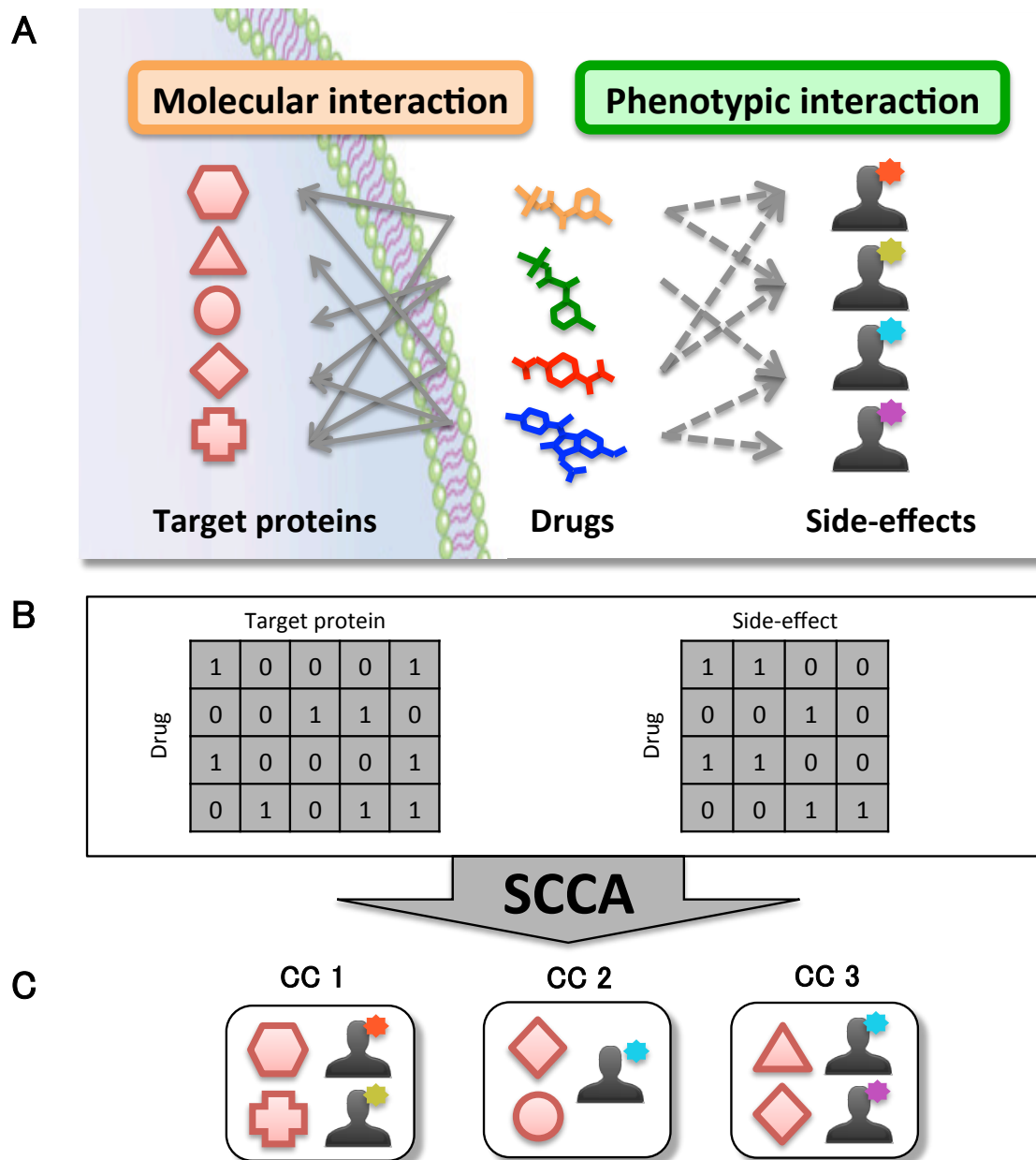


図 4 スパース正準相関分析 (SCCA) による相関成分 (CC) の抽出
 (A) 同一の薬物について、標的タンパク質との相互作用 (Molecular interaction) と、副作用キーワードとの対応 (Phenotypic interaction) を取得する。
 (B) 行を薬物、列を標的タンパク質、あるいは、副作用キーワードとする二値行列を作成する。
 (C) SCCA を適用して、標的タンパク質と副作用キーワードの集合である相関成分 (CC) を出力する。

(式の導出)

n 個のサンプル (ここでは薬物) の各々について、 p 種の特徴 (ここでは標的タンパク質) からなる特徴ベクトル $\mathbf{x} = (x_1, \dots, x_p)^T$ と q 種の特徴 (ここでは副作用キーワード) からなる特徴ベクトル $\mathbf{y} = (y_1, \dots, y_q)^T$ を考える。データの作り方により $\mathbf{x} \in \{1, 0\}^p$, $\mathbf{y} \in \{1, 0\}^q$.

このとき、標的タンパク質の特徴の任意の線形結合による新変量 $u_i = \boldsymbol{\alpha}^T \mathbf{x}_i$ ($i = 1, \dots, n$) と、副作用キーワードの特徴の任意の線形結合による新変量 $v_i = \boldsymbol{\beta}^T \mathbf{y}_i$ ($i = 1, \dots, n$) を考える。 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T$ と $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)^T$ は重みのベクトルである。

正準相関分析では、この 2 種類のデータセットの間の正準相関係数 ρ を式 (1) のように定義し、 ρ が最大になるような重みのベクトル $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ を定めることが主題となる。

$$\rho = \frac{\sum_{i=1}^n \boldsymbol{\alpha}^T \mathbf{x}_i \cdot \boldsymbol{\beta}^T \mathbf{y}_i}{\sqrt{\sum_{i=1}^n (\boldsymbol{\alpha}^T \mathbf{x}_i)^2} \sqrt{\sum_{i=1}^n (\boldsymbol{\beta}^T \mathbf{y}_i)^2}} \quad (1)$$

ここで、 $\sum_{i=1}^n u_i = 0$, $\sum_{i=1}^n v_i = 0$.

このように定めた新変量 u_i と v_i を正準変量という。

$X = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$ で定義した $n \times p$ 行列 X と、 $Y = [\mathbf{y}_1, \dots, \mathbf{y}_n]^T$ で定義した $n \times q$ 行列 Y を用いると、上記の最大化問題は次のように書ける。

$$\max\{\boldsymbol{\alpha}^T X^T Y \boldsymbol{\beta}\} \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_2^2 = 1, \quad \|\boldsymbol{\beta}\|_2^2 = 1 \quad (2)$$

ただし、 $\|\cdot\|_2^2$ はベクトル要素の二乗の和である。

2.2.3 スパース正準相関分析

正準相関分析により定められる重みベクトル $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ の大部分の要素は 0 でない。本研究では、重みベクトル $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ の要素のうち正の値をとるものを取得し、その要素に該当する標的タンパク質と副作用キーワードの集合を相関成分として抽出するが、標準の CCA では多くの標的タンパク質や副作用キーワードを相関成分に抽出することになり、結果的に相関成分の特徴付けを行うのに困難を生じる。実際の応用には、 p 種の特徴と q 種の特徴のそれぞれで、正準相関係数の最大化に顕著に寄与しているものだけに高い重みを定め、それ以外の要素には重み 0 を定めるような「疎」な重みベクトル $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ を見つけることが望ましい。このような問題を解決するために $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ にスパース性を課す手法が開発された [70]。この新しい手法はスパース正準相関分析 (sparse canonical correlation analysis; SCCA) と呼ばれる。これに比して、本研究では以降、従来の正準相関分析を ordinary canonical correlation analysis (OCCA) と呼ぶ事にする。

SCCA では重みベクトル α, β の定め方に新たに式 (3) のような条件を課すことにより、ほとんどの要素には 0 の重みが与えられる。

(式の導出とアルゴリズム)

α と β にスパース性を課すために、以下のように L_1 正則化の条件を加えて、上記の最大化問題を考える。

$$\begin{aligned} & \max\{\alpha^T X^T Y \beta\} \quad \text{s.t.} \\ & \|\alpha\|_2^2 = 1, \quad \|\beta\|_2^2 = 1, \quad \|\alpha\|_1 \leq c_1 \sqrt{p}, \quad \|\beta\|_1 \leq c_2 \sqrt{q} \end{aligned} \quad (3)$$

ここで $\|\cdot\|_1$ は L_1 ノルム (ベクトル要素の絶対値の和) であり、 c_1 と c_2 ($0 < c_1 \leq 1$, $0 < c_2 \leq 1$) はスパース性を調整するパラメータである。式 (3) は、行列 $Z = X^T Y$ のペナルティ付き行列分解 (penalized matrix decomposition) に帰着される [70]。

この行列分解の解は複数存在しうるため、反復的に出力するアルゴリズムを用いた。すなわち、

$$Z^{(k+1)} \leftarrow Z^{(k)} - d_k \alpha_k \beta_k^T, \quad Z^{(1)} = X^T Y \quad (4)$$

ここで $Z^{(k)}$ は k 回目の入力である。最終的な出力として m 個の重みベクトル $(\alpha_1, \beta_1), \dots, (\alpha_m, \beta_m)$ を得る。

図 5 は標的タンパク質と副作用キーワードに与えられた重みを OCCA と SCCA で比較している。OCCA ではほとんどの標的タンパク質と副作用キーワードに非ゼロの重みを与えているのに比べ、SCCA では限られた数の標的タンパク質と副作用キーワードにのみ正の重みを与えられているのが分かる。尚、正準相関分析では m 個の相関成分 (CC) を得ることができるが、図 5 の例では上位 8 個の CC のみを表示している。

ここで、副作用が未知の薬物について、標的タンパク質プロファイル \mathbf{x}_{new} から潜在的な副作用プロファイル \mathbf{y}_{new} を予測することを考える。予測スコアは、上記で定めた $\{\alpha_k\}_{k=1}^m$ および $\{\beta_k\}_{k=1}^m$ を用いて、次のように定義した [52]。

$$\mathbf{y}_{new} = \sum_{k=1}^m \beta_k \rho_k \alpha_k^T \mathbf{x}_{new} = B \Lambda A^T \mathbf{x}_{new} \quad (5)$$

ただし、 $A = [\alpha_1, \dots, \alpha_m]$, $B = [\beta_1, \dots, \beta_m]$ で、 Λ は正準相関係数を対角要素とする対角行列である。

\mathbf{y}_{new} の j 番目の要素が高いスコアを持つ時、この薬物は j 番目 ($j = 1, 2, \dots, q$) の副作用を有すると予測される。

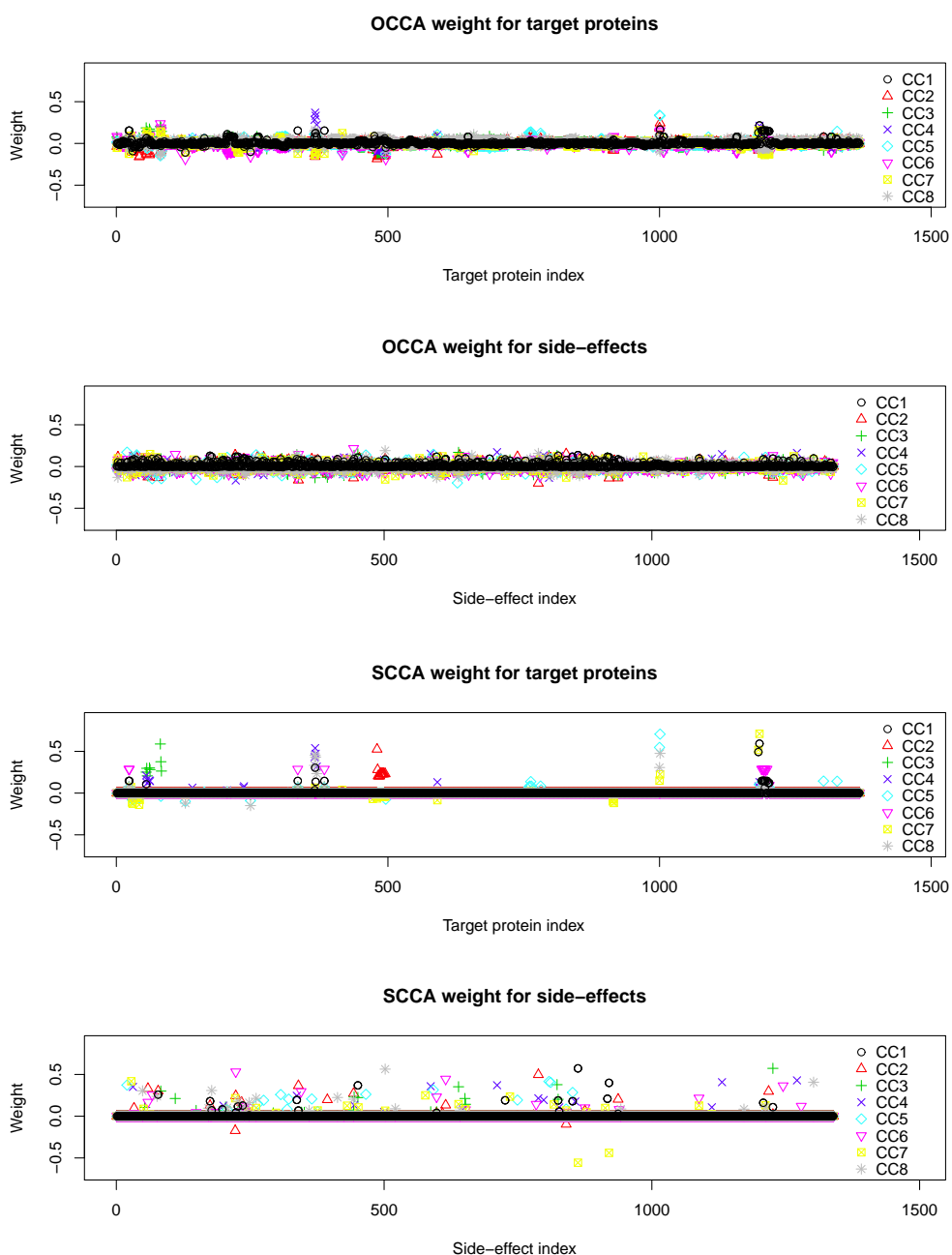


図5 OCCA と SCCA の出力の比較

各プロットの横軸は重みベクトル α のインデックス (標的タンパク質に対応) あるいは重みベクトル β のインデックス (副作用キーワードに対応) を示し、縦軸は重みの数値を示す。図では例として CC1 から CC8 までの 8 つの CC のみを表示している。1 段目。OCCA を適用したときの標的タンパク質の重み (α の値)、2 段目。OCCA を適用したときの副作用キーワードの重み (β の値)、3 段目。SCCA を適用したときの標的タンパク質の重み (α の値)、4 段目。SCCA を適用したときの副作用キーワードの重み (β の値)。

2.2.4 予測精度の検証

予測精度の検証は 5 分割交差検定を用いて行った。まず 658 個の薬物の集合をランダムに 5 分割し、その 5 分の 4 に相当する 526 個の薬物の標的タンパク質プロファイルと副作用プロファイルを学習セットとし、2.2.3 の式 (3) を解いた。残りの 5 分の 1 に相当する 132 個の薬物の標的タンパク質プロファイルと副作用プロファイルを検証データとし、式 (5) を用いて予測スコアを計算した。

予測精度の指標には受信者操作特性 (receiver operating characteristic; ROC) 曲線下面積 (area under the curve; AUC) と適合率-再現率曲線 (Precision-Recall; PR) 曲線下面積 (area under the precision-recall curve; AUPR) を用いた。AUC と AUPR は、5 分割されたデータセットのそれぞれで計算した。

相関成分 (CC) の数 m は 10 から 100 までの範囲で 10 刻みに設定し、それぞれの値で 5 分割交差検定を行い、ROC 曲線下面積が最大になるときの値を最適値とした。SCCA では、重みベクトル α と β のスパース性を調整するパラメータ c_1, c_2 の値を 0 から 1 までの範囲で 0.1 刻みに設定し、 m も含めて 3 つのパラメータの取りうる 100 通りの値の組み合わせで 5 分割交差検定を行い、ROC 曲線下面積が最大になるような最適値を求めた。ただし、計算を簡単にするために $c_1 = c_2$ とした。

2.2.5 エンリッチメント解析

CC に抽出された標的タンパク質の生物学的関連性を評価するために、タンパク質の二通りの機能単位 (1) 分子経路分類、(2) 分子機能分類を用いてエンリッチメント解析を行った。エンリッチメント解析では、CC を一つと、分子経路 (あるいは分子機能) を一つ固定したときに、CC に抽出された標的タンパク質の集合と、分子経路に關与しているタンパク質の集合の共通部分の数の有意性を統計的に評価する。CC に抽出された標的タンパク質のうち、統計的に有意に多くの標的タンパク質がこの分子経路に関わっていた場合、この CC は、この分子経路で「エンリッチされている」という。

タンパク質の分子経路分類、および分子機能分類には KEGG データベースと Gene Ontology (GO) データベースを使用した。KEGG PATHWAY から 112 枚のパスウェイマップと KEGG BRITE から 105 個の分子機能分類用語を取得した。GO biological process から 751 個の分類用語と GO molecular function から 318 個の分類用語を取得した。

エンリッチメント解析の統計的検定に用いる帰無仮説には、次の超幾何分布を用いた。
(式の導出)

G_c を CC c に抽出された標的タンパク質の集合とし、 G をある分類 (KEGG パスウェイマップなど) に含まれるタンパク質の集合とする。 $r = |G_c|$, $k = |G|$, $z = |G_c \cap G|$ とし、 l をデータセットに含まれたタンパク質の総数とする。帰無分布には超幾何分布を仮定すると、 G と G_c の共通部分の大きさが z 以上である確率は次のように計算される。

$$p(G, G_c) = \sum_{i=z}^{\min(k,r)} \frac{\binom{k}{i} \binom{l-k}{r-i}}{\binom{l}{r}} \quad (6)$$

本研究では、得られた値を擬陽性率 (False Discovery Rate; FDR) で補正し [9]、 \log を取った値 $s(c)$ を CC c のエンリッチメントスコアとした。

$$s(c) = -\log_{10} p^{\text{FDR}}(G, G_c) \quad (7)$$

2.3 結果と考察

2.3.1 薬物の標的タンパク質と副作用キーワードの相関成分の抽出

本研究では、658 個の薬物に対して 1,368 個の標的タンパク質との相互作用の有無を表した二値行列 X 、および 1,339 個の副作用キーワードとの対応の有無を表した二値行列 Y に SCCA を適用した。このとき、SCCA の出力解である重みベクトル α , β のスパース性を制御するパラメータ c_1 , c_2 と、相関成分の数 m を 5 分割交差検定により最適化した結果、最適値は $c_1 = c_2 = 0.1$, $m = 80$ であった。これにより、標的タンパク質に関する重みベクトル α と、副作用キーワードに関する重みベクトル β を 80 通り出力した。80 通りのそれぞれから、 α ベクトルの中で正の値を取る標的タンパク質の集合と、対応する β ベクトルの中で正の値を取る副作用キーワードの集合を抽出した。さらに、重みベクトルから正準相関スコアを求め、正準相関スコアが正の値を取る薬物の集合を 80 通り抽出した。このようにして得た「薬物-標的タンパク質-副作用キーワード」の 3 つ組から成る 80 個の集合をそれぞれ相関成分 (canonical component; CC) と呼ぶことにした。80 個の CC には、合計で 298 個の標的タンパク質、428 個の副作用キーワード、および 658 個の薬物が抽出された。

80 個の CC に抽出された標的タンパク質と副作用キーワードを図 6 にネットワーク表示した。ピンクのノードは標的タンパク質を表し、緑のノードは副作用キーワードを表す。視覚的に分かりやすくするために、各相関成分に抽出された標的タンパク質のうち重みの高い順に上から 5 番目までと、副作用キーワードの集合のうち重みの高い順に上から 3 番目までを表示している。標的タンパク質と副作用キーワードの対が 80 個の CC のどれかに抽出されていた場合に、該当するノード間にエッジを引いた。特に、代表的な CC1, CC2, CC5, CC15 については、高い正準相関スコアを示した薬物に着目し、これらの「薬物 標的タンパク質 副作用」の関係性について 2.3.3 で詳細に議論する。80 個の CC の全内容は補足表 S1 (http://web.kuicr.kyoto-u.ac.jp/supp/smizutan/target-effect/summary_scca_average.txt) に示した。

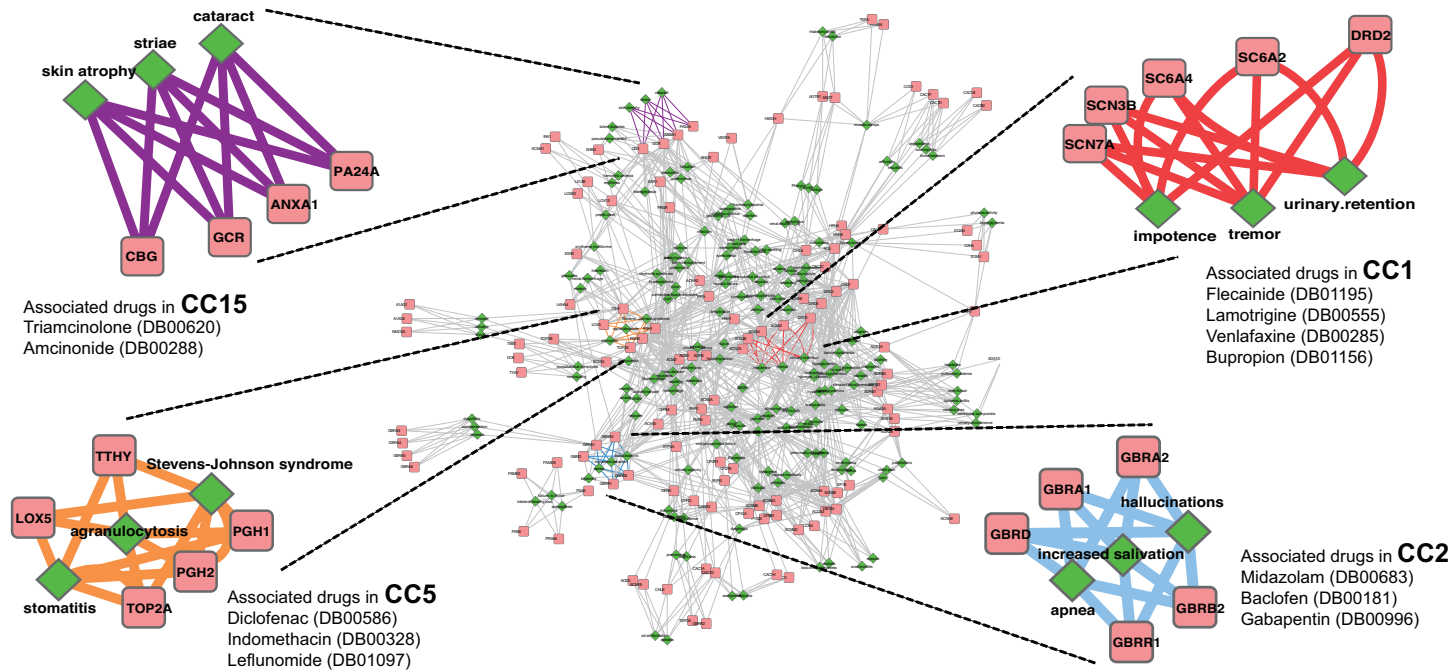


図6 80個の相関成分(CC)に抽出された標的タンパク質と副作用キーワードのネットワーク表示(Mizutaniら[45]より引用)

標的タンパク質(ピンクの長方形)と副作用(緑の菱形)が同じCCに抽出された場合にエッジを引いた。標的タンパク質は各CCで重みの高い順に上位5番目までを、副作用は上位3番目までを表示している。太線で強調したCC1(赤)、CC2(水色)、CC5(オレンジ)、CC15(紫)については、2.3.3で薬物-標的タンパク質-副作用の関係性を詳細に議論した。CC1: DRD2(ドパミンD2受容体)、SC6A2(ナトリウム依存性ノルアドレナリントランスポーター)、SC6A4(ナトリウム依存性セロトントランスポーター)、SCN(ナトリウムチャンネルタンパク質サブユニット)。CC2: GBRs(ガンマアミノ酪酸受容体サブユニット)。CC5: PGH1/2(プロスタグランジンG/Hシンターゼ1/2)、TOP2A(DNAトポイソメラーゼ2-alpha)、TTHY(トランスサイレチン)、LOX5(アラキドン酸-5-リポキシゲナーゼ)。CC15: PA24A(細胞質ホスホリパーゼA2)、ANXA1(アネキシンA1)、GCR(グルココルチコイド受容体)、CBG(コルチコステロイド結合グロブリン)。

2.3.2 相関成分に抽出された標的タンパク質の集合の生物学的検証

各 CC に抽出された標的タンパク質の集合の生物学的関連性を検証するために、タンパク質の二通りの機能単位 (1) 分子経路分類、(2) 分子機能分類によるエンリッチメント解析を行った。(1) 分子経路分類には、KEGG PATHWAY のパスウェイマップと GO biological process の分類用語を用い、(2) 分子機能分類には、KEGG BRITE による分子機能分類と GO molecular function の分類用語を用いた。エンリッチメント解析の統計値を表 3 に示す。80 個の CC に抽出された合計 298 個の標的タンパク質のうち、KEGG PATHWAY による分子経路情報は 215 個について得られ(表 3a)、GO biological process による分子経路情報は 281 個について得られた(表 3b)。KEGG BRITE と GO molecular function による分子機能情報は、298 個全ての標的タンパク質について得られた(表 3c-d)。各 CC に対し、それぞれの分類に関するエンリッチメントスコアを計算し、FDR 補正された p 値が ≤ 0.05 を充たすとき、この CC は該当する分類でエンリッチされているとした。その結果、(1) 分子経路分類では 57 個の CC が KEGG のパスウェイマップでエンリッチされ、72 個の CC が GO biological process の分類用語でエンリッチされた。(2) 分子機能分類では、75 個の CC が KEGG BRITE の分子機能分類でエンリッチされ、74 個の CC が GO molecular function の分類用語でエンリッチされた。

表 3 エンリッチメント解析の統計値 (Mizutani ら [45] より改変)

	(a)	(b)	(c)	(d)
分子経路情報・分子機能情報が得られた標的タンパク質	215	281	298	298
エンリッチメント解析に用いた分子経路・分子機能	112	751	105	318
エンリッチされた CC の数	57	72	75	74
エンリッチメントを示した分子経路・分子機能	33	93	50	75

(a) KEGG PATHWAY maps; (b) GO biological process; (c) KEGG BRITE terms; (d) GO molecular function

分子経路分類によるエンリッチメント解析

分子機能分類によるエンリッチメント解析

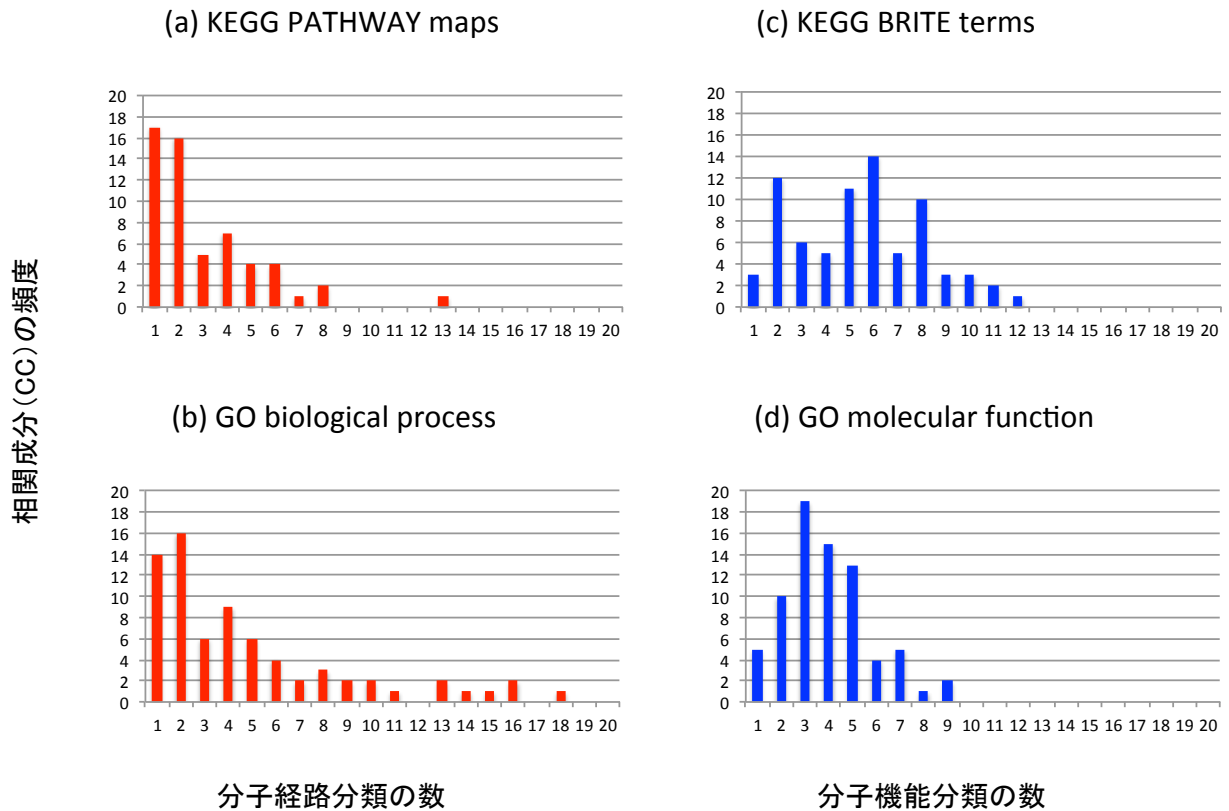


図 7 分子経路分類、あるいは、分子機能分類でのエンリッチメント解析

(赤) 分子経路分類によりエンリッチされた CC の分布

(a) KEGG PATHWAY maps を使用したエンリッチメント解析では、33 個の CC が 1 つまたは 2 つのパスウェイでエンリッチされ、残りの CC のうちの 23 個は 10 未満のパスウェイでエンリッチされた。

(b) GO biological process については、エンリッチメントを示した分類用語の数が増加するにつれて CC の数が減少した。

(青) 分子機能分類によりエンリッチされた CC の分布

(c) KEGG BRITE terms によるエンリッチメント解析では、KEGG PATHWAY maps の分布に比べて傾斜の小さい分布を示した。

(d) GO molecular function を用いたエンリッチメント解析では、分類用語の平均数が 3.95 である釣鐘型の分布を示した。

図 7 は(1)分子経路分類、(2)分子機能分類の数それぞれに対して、その数の分類でエンリッチだった相関成分(CC)数の分布を表している。分子経路分類によるエンリッチメント解析では分子経路の数が増加するにつれて CC の数は極端に減少している。特に KEGG PATHWAY のパスウェイマップを用いたエンリッチメント解析では、エンリッチメントを示した 57 個の CC のうち 33 個が 1 つまたは 2 つの分子経路でのみエンリッチされ、残りの 24 個の CC のうち 23 個は 10 未満の分子経路でエンリッチされた(図 7a)。この傾向は、各 CC の大部分の標的タンパク質が同じ分子経路で働いていることを示唆している。対照的に、KEGG BRITE による分子機能分類を用いたエンリッチメント解析では、CC は様々な数の分子機能分類に分布しているのが分かる(図 7c)。すなわち、CC の大部分はいくつかの異なる分子機能を持つ標的タンパク質から構成されている傾向が読み取れる。GO biological process (分子経路分類に相当)(図 7b)と molecular function (分子機能分類)(図 7d)を用いたエンリッチメント解析でも同様の分布の違いが示された。

しかしながら、KEGG BRITE の分子機能分類やと GO の分類用語は互いに親子関係にあるような階層構造を持つため、そのような分類用語への二重カウントにより、CC の数が図 7(c-d)のような分布を示した可能性が否定できない。そこで、分子経路でエンリッチされた標的タンパク質について、KEGG BRITE の“Ion Channels”, “G Protein-Coupled Receptors”, “Enzymes”などのより大きな分子機能カテゴリを用いてさらなる検証を行った。その結果、各 CC で分子経路エンリッチメントを示した標的タンパク質は、この分子機能カテゴリによる分類でも、異なる複数の分子機能を持つことが追認された(図 8)。この結果は 80 個の多くの CC で示された。

CC distribution over KEGG BRITE categories

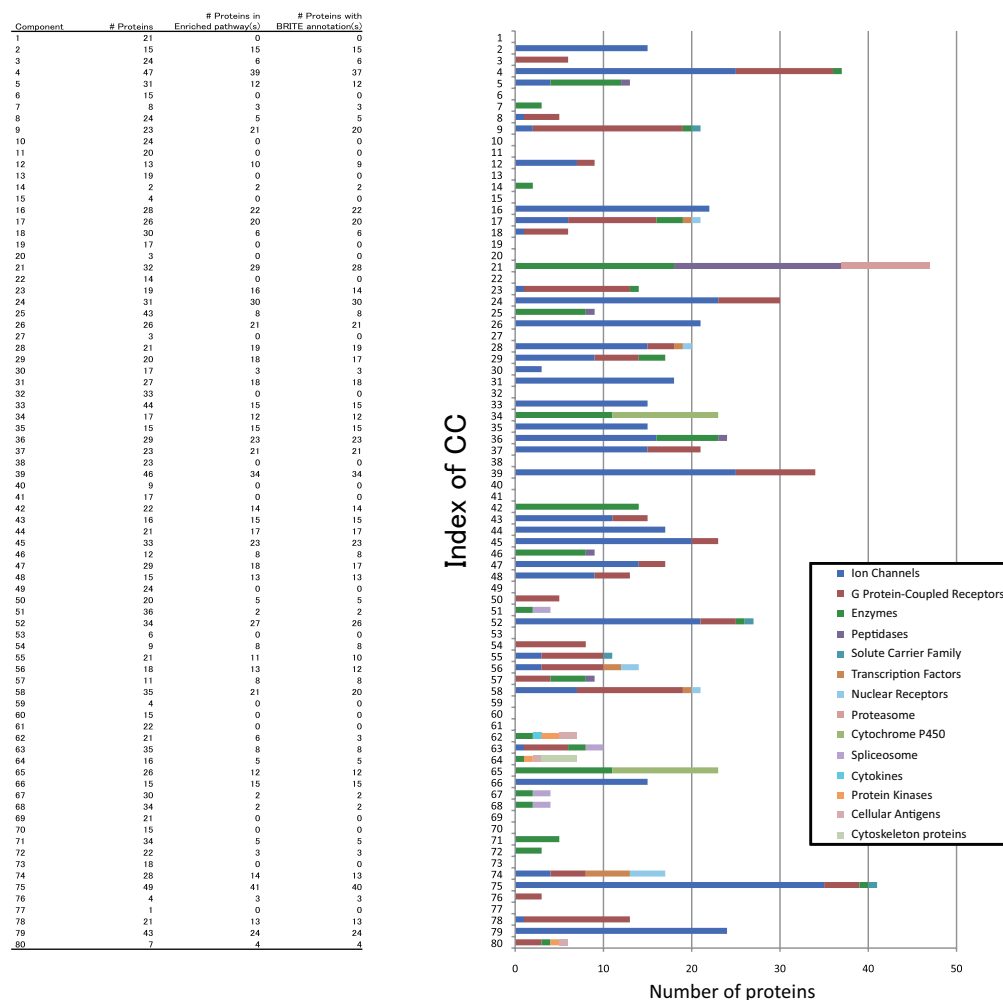


図 8 分子経路でエンリッチされた標的タンパク質の分子機能カテゴリ別分類 (Mizutani ら [45] より引用)

KEGG pathways によりエンリッチされた標的タンパク質は、KEGG BRITE の大分類では 14 の分子機能カテゴリに分類された (凡例)。

(左表) 列 1: CC 番号; 列 2: CC に抽出された標的タンパク質の数; 列 3: 分子経路でエンリッチされた標的タンパク質の数; 列 4: KEGG BRITE の分子機能カテゴリ情報の得られた標的タンパク質の数。

(右ヒストグラム) 14 種類の分子機能カテゴリごとの標的タンパク質の数。複数のカテゴリに分類された標的タンパク質は二重に数えられていることに注意。全体的な傾向として、パスウェイでのエンリッチメントを示した標的タンパク質は、分子機能カテゴリ別で見ると、“Ion Channels”、“G Protein-Coupled Receptors”、“Enzymes” など異なる分子機能カテゴリに分類されている。

次に、エンリッチメントを示した KEGG のパスウェイマップを表 4 に示した。該当する CC の多い順に上位 10 個を載せている。80 個の CC に関する全てのエンリッチメントパスウェイとエンリッチメントスコアは補足表 S2 (<http://web.kuicr.kyoto-u.ac.jp/supp/smizutan/target-effect/TableS2.pdf>) に示した。2 つ以上のパスウェイでエンリッチされた 40 個の CC のうち 14 個は、*Calcium signaling pathway* (カルシウムシグナル伝達経路, map04020)、*MAPK signaling pathway* (MAPK シグナル伝達経路, map04010)、*Cardiac muscle contraction* (心筋の収縮, map04260) のうちの 2 つから 3 つでエンリッチされていた。カルシウムイオンは心筋の収縮の制御に関わることは良く知られているが、興味深いことに、MAP キナーゼ活性がカルシウム非依存的に平滑筋の収縮に関係しているとの報告もある [15]。このようにパスウェイ間の機能の関係を考慮すると、複数の経路でエンリッチされた CC は、パスウェイ間のクロストークを表していると解釈できる場合がある。

ここで留意すべきことは、KEGG のパスウェイマップの中には、パスウェイ全体として、必ずしもタンパク質間の情報の伝達を表現していないものもあることである。例えば、*Neuroactive ligand-receptor interaction pathway* (神経刺激性リガンドと受容体の相互作用, map04080) のように、GPCR やチャネルとそのリガンドの対を複数表示してあるだけのパスウェイマップもある。このようなパスウェイマップでエンリッチされたとしても、標的タンパク質が全て互いに相互作用しているとは言えない。しかしながら、map04080 でエンリッチされた 19 個の CC のうち 18 個は、他にもエンリッチパスウェイを持っていた。

これらの結果から、各 CC に抽出された標的タンパク質は、以下の 2 つの意味で大変興味深いグループを形成していると結論付けられる。第一に、その多くが少数の共通な分子経路に関与している。第二に、そのような標的タンパク質は異なるタンパク質ファミリーに属する。この 2 つの傾向は、80 個の CC の多くに見られた。多くの場合、同じ薬物が、異なるタンパク質ファミリーに作用することは考えにくいいため、次のような解釈が可能となる。すなわち、標的タンパク質同士が異なるファミリーに属し、そのため別々の薬物により活性制御を受ける場合でも、その影響は共通の分子経路を介して同じような効果をもたらす。ここでいう効果とは、副作用の発現のことである。本研究では、副作用プロファイルの情報を反映して標的タンパク質のグループを CC に抽出している理由から、このような薬物と標的タンパク質の関係は、標的タンパク質情報のみから得られる関係よりも、副作用の発生により特化した関係であると考えられる。

表 4 多くの相関成分 (CC) でエンリッチメントを示した KEGG パスウェイマップ
(Mizutani ら [45] より引用)

ID	KEGG PATHWAY maps
map04080	Neuroactive ligand-receptor interaction
map04020	Calcium signaling pathway
map04728	Dopaminergic synapse
map04010	MAPK signaling pathway
map04260	Cardiac muscle contraction
map04727	GABAergic synapse
map04970	Salivary secretion
map04725	Cholinergic synapse
map00590	Arachidonic acid metabolism
map04270	Vascular smooth muscle contraction

80 個の CC でエンリッチメントを示したパスウェイとエンリッチメントスコアは補足表 S2 <http://web.kuicr.kyoto-u.ac.jp/supp/smizutan/target-effect/TableS2.pdf> に示した。

2.3.3 抽出された相関成分の例

2.3.2の結果をうけて、特徴的な傾向を示すCCが具体的にどのような分子機能と分子経路で説明できるかを調べた。図6でハイライトしたCC1, CC2, CC5, CC15を例に挙げて述べる。CCに抽出された薬物を詳細に調べたところ、標的タンパク質は異なるが、標的タンパク質の機能的類似あるいは生化学的経路上での相互作用のため副作用が起こるケースを、期待されていたとおり検出できた。

CC1(図6 CC1)には電位依存的セロトニントランスポーター(sodium-dependent serotonin transporter; SC6A4)および電位依存的ノルアドレナリントランスポーター(Sodium-dependent noradrenaline transporter; SC6A2)が抽出された。これらの神経伝達物質輸送体は、神経細胞のシナプスでのセロトニンとノルアドレナリンの放出および再取り込みを担っている。CC1ではまた、ナトリウムチャンネル(sodium channels; SCN3B, SCN7A)も抽出された。神経伝達物質の取り込みはナトリウムチャンネルによるナトリウムイオンの輸送と共役しているため、ベンラファキシン(Venlafaxine)またはブプロピオン(Bupropion)(図9a)による神経伝達物質輸送体の阻害と、フレカイニド(Flecainide)またはラモトリギン(Lamotrigine)(図9b)によるナトリウムチャンネルの遮断は、どちらも神経伝達物質の輸送を抑制することで共通の副作用を発生されると解釈することができる。セロトニンおよびノルアドレナリン輸送体を対象としたチャンネル遮断薬は非常に多様な化学構造を持つので、化学構造の類似性による副作用予測は期待できない。CC1には、ドパミン受容体(dopamine receptors; DRD2)も抽出された。ドパミン受容体はGPCRファミリーに属し、認知、意欲、微細な運動制御を含む多くの神経学的な機能に関与している。ドパミンはノルアドレナリンの前駆体であり、両者は化学構造的に類似しているため(図9c)、これらが共通の標的タンパク質を介して副作用を共有することは化学構造からでも推測できる。このケースを説明する薬物の例としてベンラファキシン(Venlafaxine)が挙げられる。ベンラファキシンは、上記の輸送体の他にドパミン受容体にも結合する。

CC2の標的タンパク質には、 γ -アミノ酪酸(γ -aminobutyric acid; GABA)受容体(図6のGBRA1, GBRA2, GBRB2, GBRD, GBRR1)が抽出された。GABA受容体には、イオンチャンネル内蔵型とGPCR型の分類がある。GABA_A受容体は、リガンドであるGABAの結合によりCl⁻チャンネル開口を制御するイオンチャンネル内蔵型受容体である。GABA_B受容体もGABAをリガンドとするが、Gタンパク質と共役し下流のシグナル伝達を活性化するGPCR型受容体である。CC2で高い正準相関スコアを示した薬物のほとんどは、ミダゾラム(Midazolam)(図9d)のようにGABA_A受容体

に結合する抗不安薬や麻酔補助剤であったが、バクロフェン (Baclofen) (図 9d) のような GPCR 型の GABA_B 受容体 (GBRA1,2) にも結合する筋弛緩薬も含まれていた (補足資料の Softwares and Data (<http://web.kuicr.kyoto-u.ac.jp/supp/smizutan/target-effect/bio2mat.txt>) を参照)。興味深いことに、ガバペンチン (Gabapentin) (図 9e) は GABA 受容体ではなく、電位依存型カルシウムチャネルへの阻害作用を持つ。この薬物はシナプスでの GABA 濃度を高めることが知られており [53]、GABA 受容体へ間接的に働くことで抗不安の効果を持つと考えられる。

CC5 の標的タンパク質には、プロスタグランジン G/H シンターゼ (prostaglandin G/H synthase 1,2; PGH1,2) (cyclooxygenase; COX に同じ) やアラキドン酸-5-リポキシゲナーゼ (arachidonate 5-lipoxygenase; LOX5) など炎症性の生理活性脂質の合成経路で働く酵素が抽出された。CC5 で高い正準相関スコアを示した薬物のうち、ジクロフェナク (Diclofenac) とインドメタシン (Indomethacin) (図 9f) はプロスタグランジン G/H シンターゼを阻害し、レフルノミド (Leflunomide) はアラキドン酸-5-リポキシゲナーゼに結合する。これらの薬物は、異なるタンパク質を標的とするが、生理活性脂質合成経路を阻害することで同等の抗炎症作用を発揮するため、類似した副作用をもたらすのではないかと考えることができる。

CC15 では、グルココルチコイド受容体 (glucocorticoid receptor; GCR) (核内受容体) やコルチコステロイド結合グロブリン (corticosteroid-binding globulin; CBG) などのグルココルチコイドシグナル伝達経路に關与する標的タンパク質が抽出された。グルココルチコイドはステロイドホルモンの一種であり、抗炎症性のフィードバック機構の一部を通じて免疫システムを下方制御する。この経路では、グルココルチコイドの結合によって活性化された細胞質グルココルチコイド受容体は核内移行し、アネキシン 1 (annexin A1; ANXA1) などの抗炎症タンパク質の上方制御を担う。CC15 で高い正準相関スコアを示した薬物は、トリアムシノロン (Triamcinolone)、アムシノニド (Amcinonide) (図 9h) などのグルココルチコイド受容体に作用するステロイド分子があり、これらの薬物の共通の副作用発現の機序として、グルココルチコイドシグナル伝達経路が関わっている事が推定される。

CC5 と CC15 はどちらも炎症作用に関わるタンパク質とそれを制御する薬剤によって構成されているが、この二つは作用機序を異にするため CC5 と CC15 では抽出された副作用が異なる。

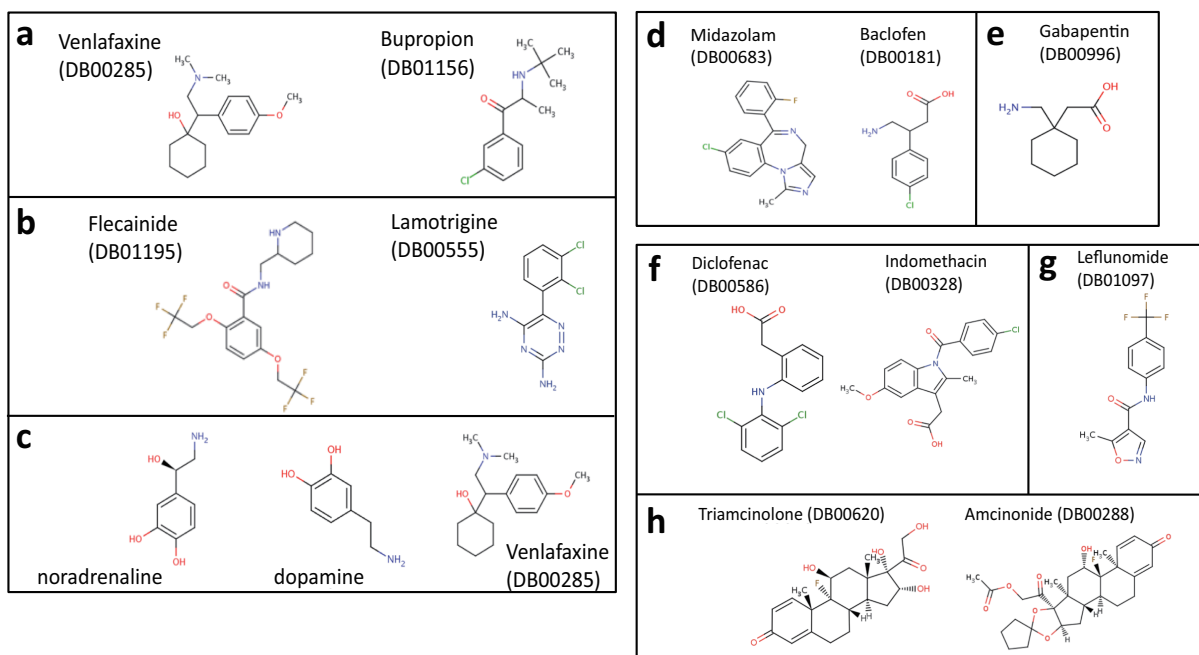


図9 高い正準相関スコアを示した薬物の例 (Mizutani ら [45] より引用)

(a-c. CC1) ベンラファキシン (Venlafaxine) およびブプロピオン (Bupropion) は神経伝達物質輸送体に結合する。フレカイニド (Flecainide) およびラモトリジン (Lamotrigine) はナトリウムチャンネルに結合する。ノルアドレナリン (noradrenaline) はノルアドレナリントランスポーターの内在性リガンドである。ドパミン (dopamine) はドパミン受容体の内在性リガンドである。ベンラファキシンはノルアドレナリン輸送体およびドパミン受容体に結合する。

(d-e. CC2) ミダゾラム (Midazolam) は $GABA_A$ 受容体に、バクロフェン (Baclofen) は $GABA_A$ および $GABA_B$ 受容体に作用する。ガバペンチン (Gabapentin) は、シナプスでの GABA 濃度を増加させることによって間接的に GABA 受容体の活性を制御する。

(f-g. CC5) 性抗炎症薬 (NSAID) であるジクロフェナク (Diclofenac) とインドメタシン (Indomethacin) はプロスタグランジン G/ H シンターゼ (prostaglandin G/H synthase 1,2; PGH1,2) (cyclooxygenase; COX に同じ) に結合し、レフルノミド (Leflunomide) はアラキドン酸 5-リポキシゲナーゼ (arachidonate 5-lipoxygenase; LOX5) に結合する。

(h; CC15) トリアムシノロン (Triamcinolone) とアムシノニド (Amcinonide) は、グルココルチコイドシグナル伝達経路のタンパク質に結合する非ステロイド性抗炎症薬である。

括弧内は DrugBank ID である。

2.3.4 副作用予測への応用

薬物には、標的となるタンパク質がよく知られている一方、副作用が未知であるものが多くある。未知の副作用を予測するひとつの方法として、標的タンパク質プロファイルから副作用を予測するアプローチが考えられる。我々は、上記の解析法を応用して、標的タンパク質プロファイルに SCCA を適用することにより未知の副作用を予測した。

2.3.4.1 予測精度の検証 まず、提案手法の予測精度を検証した。これまでの研究では、Atias と Sharan が OCCA を用いて [4]、Pauwels らが SCCA を用いて [52] 薬物の化学構造プロファイルから副作用を予測する方法を提案している。本研究で提案した予測方法の精度を検証するにあたり、これらの既存の方法を比較対象とした（表 5、図 10、図 11）。スパース正準相関分析では特徴抽出の際にスパース性を調整するパラメーターを最適化した。

表 5 予測精度の比較 (Mizutani ら [45] より引用)

Method	AUC \pm S.D.	AUPR \pm S.D.
Chemical structure based approach		
Random	0.5000 \pm 0.0000	0.0556 \pm 0.0000
OCCA (Atias と Sharan の手法)	0.8355 \pm 0.0010	0.3753 \pm 0.0016
SCCA (Pauwels らの手法)	0.8708 \pm 0.0007	0.3766 \pm 0.0030
Target protein based approach		
Random	0.5000 \pm 0.0000	0.0556 \pm 0.0000
OCCA	0.8850 \pm 0.0007	0.4067 \pm 0.0006
SCCA (本研究の提案手法)	0.8895 \pm 0.0002	0.4103 \pm 0.0018

表 5 と図 10-11 は 4 つのアプローチ間での予測精度を比較している。標的タンパク質プロファイルに基づいて副作用予測を行った場合と、化学構造プロファイルに基づいて副作用予測を行った場合の予測精度を比較した。既存の 2 つの手法と比較するために SCCA による予測手法の他に OCCA での精度も確認した。予測精度の指標には、ROC 曲線下面積 (AUC) と Precision-Recall 曲線下面積 (AUPR) を用いた。AUC は予測がランダムである場合は 0.5 となる。4 つのどのアプローチも高い予測精度を出す。SCCA はタンパク質プロファイル、化合物フィンガープリントのどちらを用いた場合でも OCCA よりやや優れていた。またデータにタンパク質プロファイルを用いた場合は、

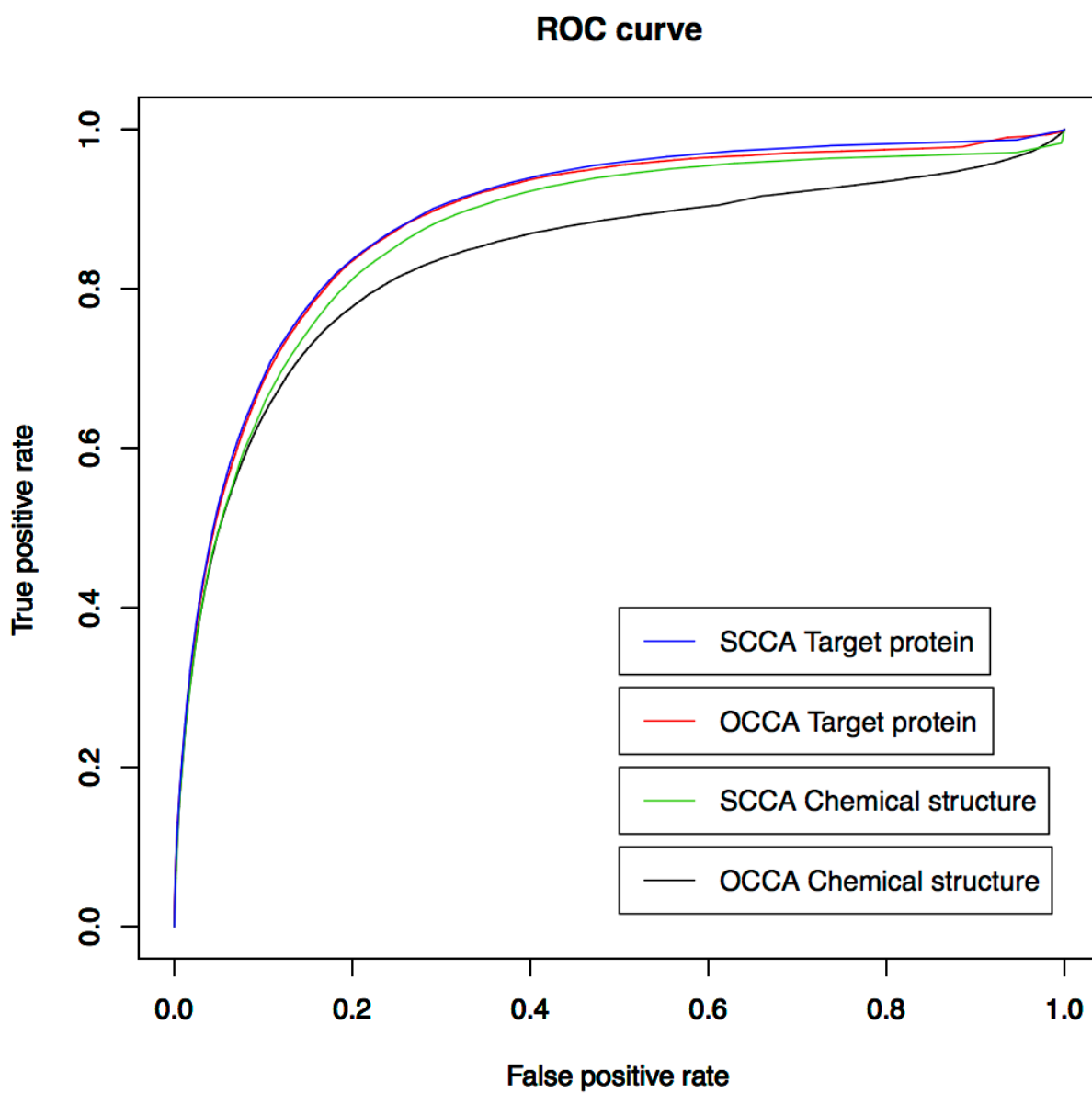


図 10 ROC 曲線を用いた予測精度の比較 (Mizutani ら [45] より引用)

Precision-Recall curve

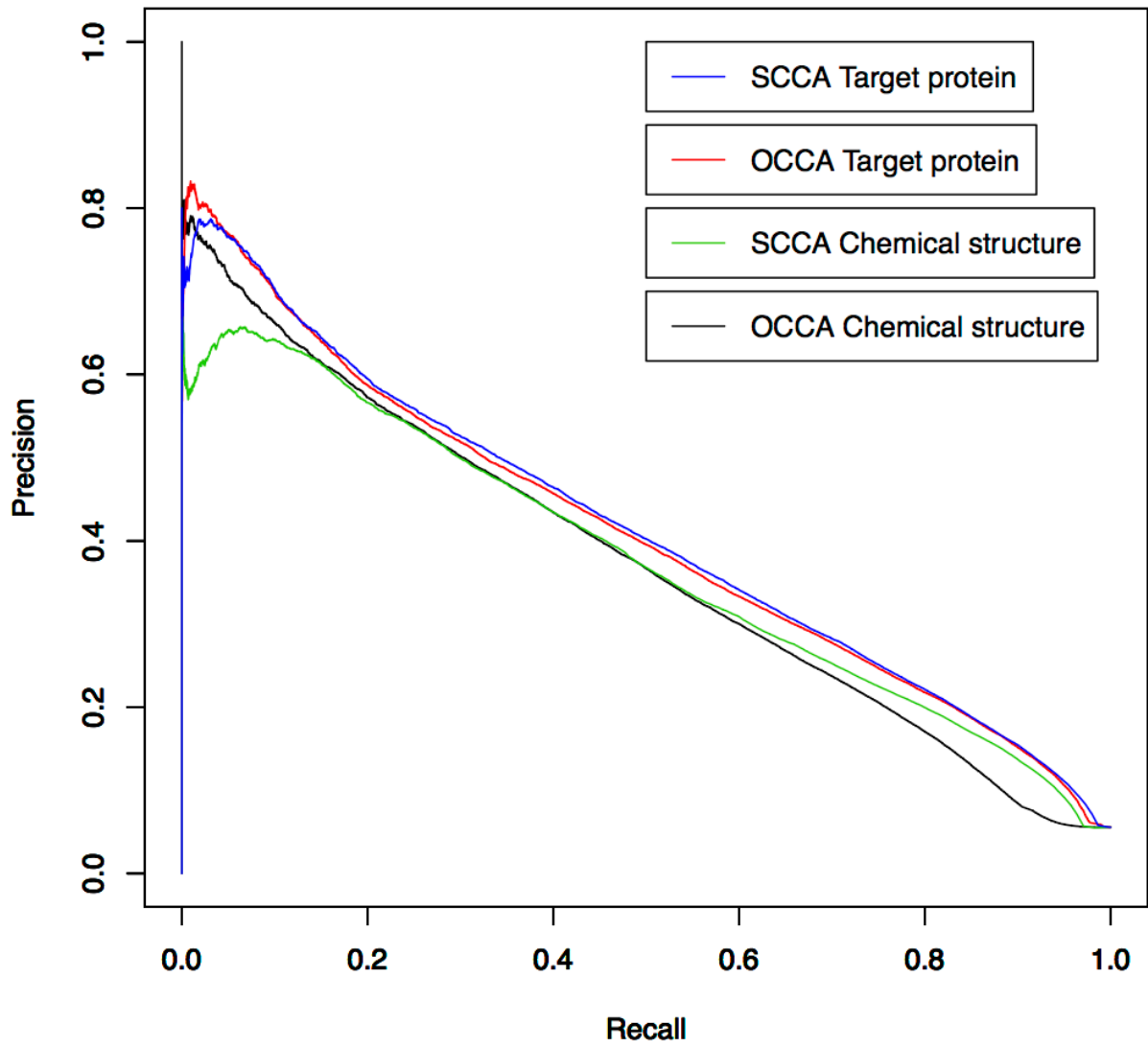


図 11 Precision-Recall 曲線を用いた予測精度の比較 (Mizutani ら [45] より引用)

OCCA, SCCA のどちらの手法でも、化合物フィンガープリントを用いた場合の精度を上回った。この検証結果から、標的タンパク質情報に SCCA を適用するアプローチは副作用予測に有用であることが示された。

2.3.4.2 副作用の予測 DrugBank データベースにより標的タンパク質情報が提供されている 1,388 個の薬物のうち、SIDER データベースに副作用情報が登録されていない薬物が 730 個あった。本研究では、そのような薬物の対して、標的タンパク質プロファイルに基づき、提案した SCCA モデルベースの予測器を用いて、潜在的な副作用を予測した。ここでは、前述の 2.3.1 で 80 個の相関成分を同定した際に用いた 658 個の薬物の標的タンパク質プロファイルデータと副作用プロファイルデータを学習セットとして使用し、SCCA モデルベースの予測器を学習させた。予測された副作用のうち、高いスコアで予測された副作用に焦点を当てて文献調査を行い、その妥当性を検証した (表 6)。すべての予測結果は、補足表 S3 (http://web.kuicr.kyoto-u.ac.jp/supp/smizutan/target-effect/newpredpairmat_bio.txt) に示した。

最も高いスコアで予測された薬物であるシンナリジン (Cinnarizine) は酔い止めに使用される抗ヒスタミン薬である。この薬物はヒスタミン H1 受容体に結合し、嘔吐を予防することから乗り物酔いに有効性を持つと考えられている。我々の方法から、シンナリジンは副作用に「振戦^{*8}」を持つことが予測され、その妥当性が文献 [23] により確認された。この副作用は、シンナリジンが電位依存性カルシウムチャンネルに結合し筋肉の収縮に関与することで起こると考えられる。他に予測された副作用として「便秘」があるが、これは FDA の有害作用報告 6127929-0 によって指摘され、「口渇」も文献 [24] によって確認されている。

予測の第 2 位は麻酔薬であるベンゾカイン (Benzocaine) であった。ベンゾカインは電位依存性ナトリウムチャンネル阻害剤であり、神経線維に沿ったパルスの伝達を妨げる作用がある。その副作用として「複視」が起こることが予測され、文献 [30] で確認された。他にも、「失神」は文献 [65] によって確認された。

予測結果の第 4 位は高血圧症治療薬であるベプリジル (Bepridil) の「振戦」である。「振戦」はベプリジルに最も良く観られる副作用のひとつであり、患者の 5 % がこの副作用を報告している [69]。

最後に、抗精神病薬であるプロマジン (Promazine) が「頻脈」を引き起こすことや [3]、トップ 20 位以内には入らなかったため表には載せていないが、高血圧の管理に使用され

*8 振戦 (tremor): 筋肉の不随意的な収縮と弛緩の繰り返し。

るカルシウムチャネル遮断薬であるニソルジピン (Nisoldipine) が「複視」を引き起こすことも文献で確認された [49]。

表 6 標的タンパク質プロファイルに基づき SCCA を用いて予測された副作用 (上位 20)

ランク	薬物名	DrugBank ID	副作用	予測スコア	文献
1	Cinnarizine	DB00568	振戦	1.176339	[23]
2	Benzocaine	DB01086	複視	1.163882	[30]
3	Cinnarizine	DB00568	便秘	1.14319	FDA 有害作用報告 6127929-0
4	Bepidil	DB01244	振戦	1.046472	[69]
5	Cinnarizine	DB00568	傾眠	0.99626	—
6	Cinnarizine	DB00568	口渇	0.961833	[24]
7	Cinnarizine	DB00568	血管浮腫	0.955471	—
8	Cinnarizine	DB00568	不眠	0.950757	—
9	Benzocaine	DB01086	吐気	0.947105	—
10	Alprenolol	DB00866	眩暈	0.943918	—
11	Benzocaine	DB01086	下痢	0.937409	—
12	Nisoldipine	DB00401	振戦	0.933258	—
13	Nitrendipine	DB01054	振戦	0.933258	—
14	Lercanidipine	DB00528	振戦	0.933258	—
15	Benzocaine	DB01086	失神	0.928905	[65]
16	Promazine	DB00420	頻脈	0.926875	[3]
17	Promazine	DB00420	傾眠	0.924103	—
18	Benzocaine	DB01086	嘔吐	0.922091	—
19	Bepidil	DB01244	神経過敏	0.918828	—
20	Cinnarizine	DB00568	神経過敏	0.918615	—

文献で確認が取れた予測結果には、該当する文献番号を示す。

2.4 まとめと今後の展望

研究 I では、標的タンパク質との相互作用情報を用いることで、副作用の発現機序に迫ることができるのではないかと考え、薬物と標的タンパク質との相互作用情報と副作用を関連付ける方法を提案し、分子経路情報を用いていくつかの発現機序について議論した。

ほとんどの創薬プロジェクトでは、注目している疾患の原因となる治療標的を探索し、それらが関与する分子経路を同定する。薬物の副作用については、非臨床試験での実験動物を使った毒性試験、およびヒトを対象とする臨床試験（第 I 相および第 III 相）で安全性の検証を行う。実際、hERG など副作用を引き起こすことが良く知られているタンパク質についてはアッセイ系が確立されているが、費用がかかるのは事実である。近年、薬物の作用は単体の標的分子だけでなく、その周辺の分子ネットワークを通じて広く影響を及ぼすことが認識されてきたが、タンパク質同士のつながりも考慮したアッセイ系の構築は非常に難しい。そこで、公開データベースに蓄積された膨大な知識・情報をバイオインフォマティクス手法を用いて統合することにより薬理的、および、生物学的に有用な関連を抽出するアプローチが有効である。本研究では、薬物・標的タンパク質・副作用という 3 つのデータ空間から重要な関係性を抽出する解析法を提案し、標的タンパク質のネットワークの視点から副作用の発現機序を推定した。その応用として、標的タンパク質プロファイルから未知の副作用を予測した。添付文書などから得られる副作用情報は標的タンパク質情報に比べてはるかに少ないので、薬物の標的タンパク質プロファイルから副作用プロファイルを予測することは、創薬を効率化する妥当な手段であると言える。本研究で使用了手法と、本法を用いて得られた結果について、今後の展望を述べる。

2.4.1 手法に関する展望

本研究では、同一の薬物に関して標的タンパク質プロファイルと副作用プロファイルという 2 種類のデータセットの間の相関関係を解析する手法として、多変量解析法の一つである SCCA を用いて関連解析を行った。SCCA を含め、多くの関連解析の枠組みでは、データ集合から関連性の強いものを部分集合として抽出する。その利点は、抽出された部分集合を、知識情報を用いて特徴付けることにより解析対象をシステムとして理解できる点にある。薬物の標的タンパク質は、生体内では多くの場合、単体ではなく、他の生体内物質との情報のやりとりを介して機能を果たすことができる。すなわち、標的タンパク質の集合を意味のある集合に抽出し、分子機能分類や分子経路分類などの生命知識情報を用いて特徴付ける解析法が妥当であると考えられる。

一方、このような解析法の弱点は、直接の因果関係を明らかにすることが難しい点にある。Kuhnらは、本研究も含め、副作用に関する既存の研究を比較評価している（表7 [37]）。そこでは、本研究は、標的タンパク質と副作用の間での因果関係を明らかにするところへは至っていないと批判している。

Kuhnらは、SIDER2 データベース [38] と STITCH3 データベース [39] から取得した副作用情報と標的分子情報を元に、統計的に有意に多くの薬物に現れる標的タンパク質と副作用のペアを抽出している [37]。Kuhnらの方法では、フィッシャーの正確確率検定を用いているため、標的タンパク質と副作用の一対一の予測が可能である。さらに、予測結果を検証実験で裏付けることにより、標的タンパク質と副作用の因果関係を明らかにしている。Benderらが用いたマルチカテゴリベイズモデル [8] も一対一の関係を予測することに注目している。しかしながら、Kuhnの解析では分子経路の視点に立った副作用発現機序の検証はしておらず、この意味においては、本研究はKuhnらの研究と目指すところを異にする。また、生体をシステムと捉え、薬物の作用機序をタンパク質ネットワークへの介入と捉える見方においては、本研究で得られた情報が有効であるといえるだろう。

Kuhnらは、計算機科学的に予測した標的タンパク質と副作用のペアに対して、文献検索とマウスを使った *in vivo* 実験により検証を行っている。本研究で提案した解析法に関しては、外部リソースに対するベンチマークテストがされていないと批判している。本研究では、データソースは1つで、予測精度の検証はデータを学習データと検証データに分割して、交差検定を行った。外部データソースを用いて予測精度を検証することにより、予測器の汎用性が高まるため、このことは今後の課題である。また、文献検索による網羅的な検証を行っていないとの批判もある。文献の網羅的な検索は手作業では不可能であり、PubMedに蓄積された文献を計算機手法を用いて網羅的に検索するなどの方法の構築が必要である。例えば、PubMedデータベースはAPIを使って論文のタイトルや要旨に対してキーワード検索をかけることができるため、薬物と副作用キーワードのペアをクエリとした網羅的な検索が可能である。しかしながら、単純な検索では精度が低いため、要旨中の文脈を考慮したより高度な検索や手作業による検証が必要である。

表 7 副作用の研究の比較 (Kuhn ら [37] より改変)

	Mizutani <i>et al.</i> , 2012 [45] (本研究)	Khun <i>et al.</i> , 2012 [37]	Atias and Sharan, 2010 [4]	Bender <i>et al.</i> , 2007 [8]	Fliri <i>et al.</i> , 2005, 2007 [19, 20]	Scheiber <i>et al.</i> , 2009 [56]	Xie <i>et al.</i> , 2007, 2009 [71, 72]
Experimental validation	-	+	-	-	-	-	-
Validation from literature	-	+	-	+	-	+	-
Use protein structures	-	-	-	-	-	-	+
Use drug-target databases	+	+	-	+	+	+	-
Use drug-target text-mining	-	-	-	-	-	-	-
Prediction: proteins-side effects	+	+	-	(+)	-	-	(+)
Prediction: pathway-side effects	+	-	-	-	-	+	-
Prediction: drug-side effects	+	-	+	+	-	-	-
Complete set of side effects	+	+	+	+	+	+	-
Use only post-marketing side effects	-	-	-	-	-	-	-
Complete set of targets	+	+	(+)	-	+	+	-
Complete set of drugs	+	+	+	+	+	+	-

2.4.2 結果に関する展望

本研究では、相関成分 (CC) に抽出された標的タンパク質集合の生物学的関連性を評価するために、タンパク質の分子経路分類と分子機能分類を用いてエンリッチメント解析を行った (2.3.2)。その結果、そのような標的タンパク質の集合は、分子機能分類に比べ、分子経路分類という機能単位で、より顕著にエンリッチされていることが示された。しかしながら、薬物によるタンパク質ネットワークのより詳細な発現機序の解明には、さらに多くの情報が必要となる。ここでは、本研究で用いたアプローチの限界と拡張可能性について述べる。

2.4.2.1 分子経路情報を用いたエンリッチメント解析の利点と限界 エンリッチメント解析では、分子経路情報を用いても、標的タンパク質に対する薬理作用の種類 (作動、阻害など) や、その結果としての生理反応の方向性 (促進、抑制) を解析することは難しい。ここでは、CC9 でエンリッチメントを示した分子経路のひとつであるドパミンシナプス伝達経路を例に挙げてその問題点を明らかにする。

CC9 に抽出された標的タンパク質のうち、KEGG のパスウェイマップ *Dopaminergic Synapse* に関与するものを表 8 に示す。 *Dopaminergic synapse* でエンリッチされた標的タンパク質は、主に 3 つのタンパク質ファミリーに属する。

1 つ目はドパミン受容体 (dopamine receptor) の 5 つのサブタイプ D1(D1A), D2, D3, D4, D5(D1B) である。これらは全て GPCR ファミリーに属し、神経伝達物質であるドパミンをリガンドとするが、中枢神経系での機能は次の 2 種類に大別される。D1 様受容体 (D1, D5) は、アデニレートシクラーゼを活性化させ cAMP 産生を促進する。一方、D2 様受容体 (D2, D3, D4) は、アデニレートシクラーゼの活性を阻害し cAMP の産生を抑制する。D2 には、同一遺伝子の選択的スプライシングによるアイソフォームである D2L と D2S があり、D2L は後シナプスに発現し、前シナプスから放出されたドパミンをリガンドとし、神経細胞の興奮を起こす。D2S は前シナプスに発現し、ドパミンのシナプス放出のフィードバックとして神経細胞の興奮を抑制する。

2 つ目はナトリウム依存性ドパミン輸送体 (solute carrier family; DAT, SLC6A3) である。DAT は前シナプスに発現し、ドパミンの再取り込みによりシナプス間隙からのドパミンのクリアランスを行う。その結果、シナプスの興奮が抑えられる。DAT は D2S を介したドパミンの自己抑制を行うことも知られている。

3 つ目はモノアミン酸化酵素 B (monoamine oxidase B; MAO) である。この酵素は、前シナプスに取り込まれたドパミンを分解する役割を持つ。

表 8 KEGG パスウェイマップ *Dopaminergic synapse* に関する CC9 の標的タンパク質

タンパク質名	KEGG BRTE による分子機能分類	遺伝子名	UniProt ID
ドパミン受容体	G protein-coupled receptors	D1(D1A)	DRD1_HUMAN
		D2	DRD2_HUMAN
		D3	DRD3_HUMAN
		D4	DRD4_HUMAN
		D5(D1B)	DRD5_HUMAN
ナトリウム依存性ドパミン輸送体	Transporters	DAT(SLC6A3)	SC6A3_HUMAN
モノアミン酸化酵素 B	Enzymes	MAOB(AOFB)	AOFB_HUMAN

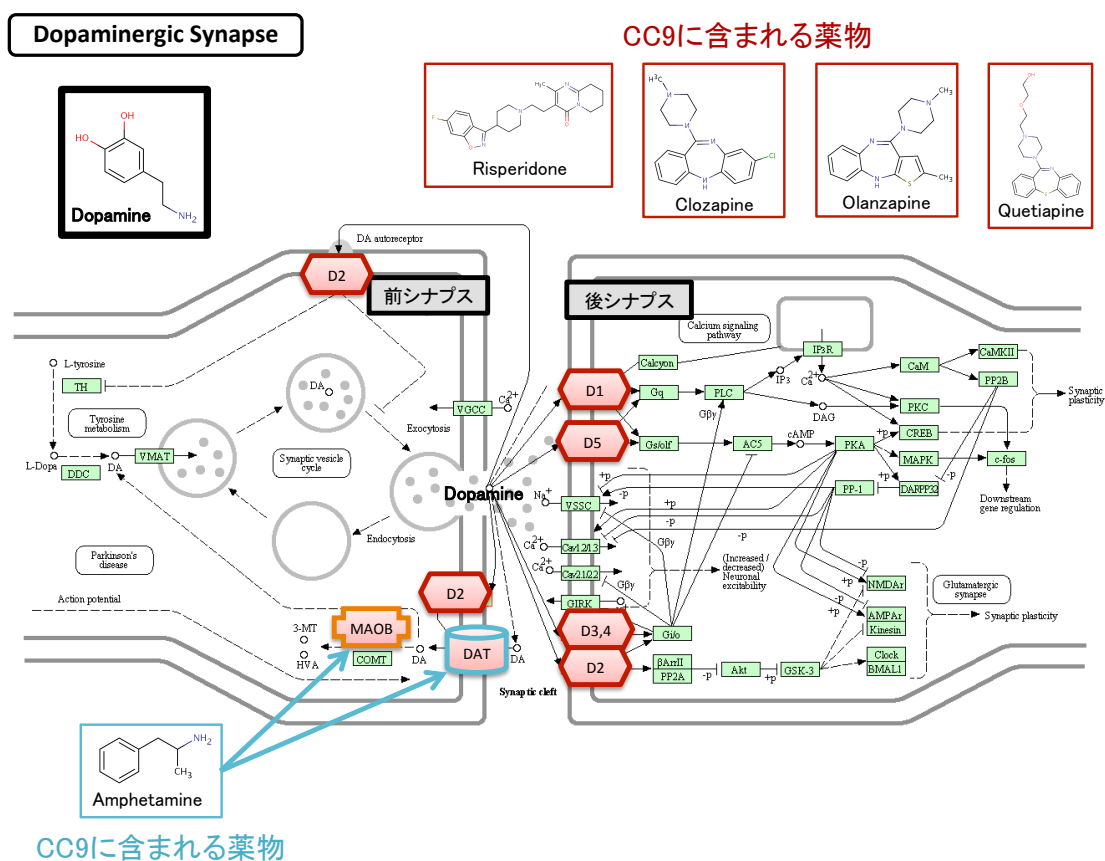


図 12 *Dopaminergic Synapse* を用いたドパミン伝達経路への薬理作用の模式図
 CC9 に抽出された標的タンパク質のうちドパミン伝達経路に関与しているものは次の3つのタンパク質ファミリーに属する。ドパミン受容体のサブタイプ (dopamine receptors; D1(D1A), D2, D3, D4, D5(D1B)) (赤枠) ナトリウム依存性ドパミン輸送体 (Sodium-dependent dopamine transporter; DAT) (水色枠) モノアミン酸化酵素 B (Amine oxidase [flavin-containing] B; MAOB) (オレンジ枠)。

表9 Dopaminergic Synapse 上の標的タンパク質に作用する CC9 の薬物

薬物名	DrugBank ID	CC9 の標的タンパク質
Risperidone	DB00734	D1(D1A), D2, D3, D5(D1B)
Clozapine	DB00363	D1(D1A), D2, D4, D5(D1B)
Olanzapine	DB00334	D1(D1A), D2, D3, D4, D5(D1B)
Quetiapine	DB01224	D1(D1A), D2, D3
Amphetamine	DB00182	DAT(SLC6A3), MAOB(AOFB)
Maprotiline	DB00934	MAOB(AOFB)* ⁹

このように、ドパミン伝達経路は、ドパミンによる神経伝達に対する正の制御（促進）と負の制御（抑制）を含む方向性のあるネットワークを形成し、全体として、ドパミンによる神経細胞の興奮とその抑制を担っている。

CC9 に抽出された薬物を表 9 に記載する。抗精神病薬であるリスペリドン（Risperidone）、クエチアピン（Quetiapine）、オランザピン（Olanzapine）は、ドパミン受容体 D2 およびセロトニン受容体 5-hydroxytryptamine（5-HT）受容体への拮抗薬であるが、ドパミン受容体のその他のサブタイプとの相互作用も知られている。特にクエチアピンは多元受容体標的抗精神病薬（multi-acting receptor targeted antipsychotic; MARTA）であり、その他にもアドレナリン受容体やヒスタミン受容体への拮抗作用も持つ。アンフェタミン（Amphetamine）は注意欠陥多動性障害*¹⁰ 治療薬で、ナトリウム依存性ドパミン輸送体（Sodium-dependent dopamine transporter; DAT）によるドパミンの再取り込みを阻害するほか、モノアミン酸化酵素 B（Amine oxidase [flavin-containing] B; MAOB）への作用も持つ。

ドパミン伝達経路の薬理作用には、ドパミンの作動と抑制という互いに逆の方向性が存在する。そのため、薬物による標的タンパク質への制御にもドパミンの作動と抑制という逆の方向性があり、薬物によってその程度も異なるはずである。しかしながら、本研究で用いたエンリッチメント解析では、制御の方向性や程度も含めた動的な作用機序の推定は難しい。さらに、薬物のいくつかはドパミン受容体だけでなく、中枢神経系に発現する様々な受容体に対する作用を持つなど、非常に複雑な制御系を形成している。そのた

*⁹ マプロチリン（Maprotiline）は MAOB（AOFB）との相互作用がデータベースに登録されていたため、CC9 に抽出されたが、その主な標的はナトリウム依存性ノルアドレナリン輸送体（Sodium-dependent noradrenaline transporter）阻害であり、モノアミンオキシダーゼへの活性制御はないとされている（DrugBank）。

*¹⁰ 注意欠陥多動性障害（attention deficit hyperactivity disorder; ADHD）：多動性、不注意、衝動性などの症状を特徴とする発達障害の一つ。）

め、ドパミン伝達経路への薬理作用や副作用の機序を理解するためには、一枚のパスウェイマップだけでなく、他のパスウェイとのクロストークなどを考慮して解析する必要がある。

2.4.2.2 標的タンパク質の組織レベルでの発現の問題 タンパク質の発現には組織特異性がある。例えば、D2 は線条体、側座核、嗅結節、大脳皮質、視床下部、黒質などに発現するが、D4 は線条体、黒質での発現が少ないなど、組織による発現の違いがある。そのため、CC9 の標的タンパク質が特定の *Dopaminergic Synapse* パスウェイでエンリッチメントを示していたとしても、それらが同じ組織で発現しているとは限らない。

薬理作用によるタンパク質制御ネットワークの改変を詳細に調べるには、細胞ごと、組織ごとの発現情報が必要となる。近年、マイクロアレイや次世代シーケンサによる遺伝子発現データが得られるようになり、細胞、組織レベルでの遺伝子発現を同定することが可能になりつつある。これらのデータを統合することにより、副作用に関係のあるタンパク質の制御ネットワークを細胞、組織別に解析することが期待できる。

2.4.2.3 薬物動態学的な副作用発現に関しての展望 薬物の作用には、主に、薬力学的な作用と薬物動態学的な作用がある。薬力学では、薬物によるタンパク質機能の促進や抑制などの薬理作用を研究する。薬理作用の種類により、作動薬 (agonist)、拮抗薬 (antagonist)、遮断薬・阻害薬 (inhibitor・blocker) などと呼ばれる。一方、薬物動態学では、生体内分子による薬物の吸収 (absorption)、分配 (distribution)、代謝 (metabolism)、排出 (excretion) を研究し、その主な解析視点を体内での輸送や濃度変化に置いている。薬物動態学では、この4つの機構に関係のあるシトクロム P450 などの薬物代謝酵素や、肝・腎輸送体を解析対象とする。

本研究では、副作用の発現機序を薬力学的な視点から解析するため、標的タンパク質と副作用とを関連づけ、分子経路情報を用いて特徴付けた。その際、薬物の代謝や排出に関わるシトクロム P450 や肝・腎輸送体などのタンパク質はデータから除外し、薬物動態学的な視点からの機序推定は行わなかった。しかしながら、肝毒性や腎毒性などの副作用の発現は、薬物の分解や体外排出などの薬物動態学的な要因によるところが大きい。副作用の機序を全般的に解明するには、薬物動態学的な視点からも解析を行う必要があるが、そのためには、解析の視点を、分子経路レベルだけでなく、臓器レベルまで広げる必要がある。また、薬物動態学的な副作用発現には薬物の濃度が深く関わっているが、そのような解析には、薬物とタンパク質の相互作用の有無だけでなく、濃度のデータも統合する必要がある。本研究で用いた関連解析では濃度情報を扱うには限界がある為、微分方程式系などを用いたシミュレーションによる動態学的な解析法が必要がある。

2.4.2.4 オフターゲットタンパク質への拡張 本研究では、薬物の主要な標的タンパク質と副作用との関係性を調べ、オフターゲットデータを積極的に採用することはしなかった。Bender ら [8] や Scheiber ら [56] の研究では、副作用の機序をオフターゲットへの作用に起因させている。副作用の発現機序を分子経路レベルで推定する上で、オフターゲットへの潜在的な薬理作用は重要である。

本研究で提案した関連解析の拡張として、オフターゲットも含めた標的タンパク質ネットワークを考慮することも可能である。ChEMBL データベースでは、薬物も含めた幅広い化合物のアッセイ実験を文献から網羅的に収集している。2014 年現在、1,411,786 個の化合物と 10,579 個の標的分子を含む 1,106,285 件の結合アッセイデータを登録している。本研究で使用した DrugBank データベースと Matador データベースの標的タンパク質との相互作用データに、ChEMBL データベースのデータを追加し、タンパク質プロファイルをオフターゲットにまで拡張することが考えられる。従って、相関成分には、主ターゲットだけでなく、オフターゲットも含まれ、両者を統合したタンパク質ネットワーク上での副作用の発現機序の推定が可能となる。しかしながら、化合物がオフターゲットに対して親和性を持つことが必ずしも薬理作用を保証するとは限らない。また、オフターゲットは、タンパク質機能が明らかでない場合もあるため、特徴付けには注意が必要である。

第3章 研究Ⅱ 薬物有害事象の疫学データを用いた副作用分類と薬剤疫学的特徴付け

3.1 序論

医薬品は、前臨床試験や臨床試験（治験）で安全性の検証を行った後に市場に出される。製薬会社は、薬剤製品の詳細な情報を添付文書に記載して薬剤を販売している。KuhnらによるSIDERデータベース [38] など、市販薬の添付文書に記載された副作用情報を登録したデータベースは、研究Ⅰでも述べたように、副作用の解析に広く利用されている。しかしながら、治験では、対象患者の数も少なく、また個人差を必ずしも反映していないので、薬物の重篤な副作用をすべて検出することは難しい。そのため、製薬会社、病院、規制当局などによる市販後調査によって継続して有効性、安全性、有用性をモニタリングしていくことが重要となる。また、近年注目されている個別化医療の設計には、患者の生理学的な違い（性別・年齢・病歴など）による副作用の種類や重篤さの違いを解明することが重要である。市販後の薬物有害反応（adverse drug reaction; ADR）の記録を大規模に収集した自発報告システム（spontaneous reporting system; SRS）のデータを解析することにより、開発段階では検出されにくい副作用を検出することができる。さらに、自発報告システムのデータには、患者の生理学的背景などの情報が付随している場合があるので、副作用発現の機序をより詳細に解析することが可能であると期待される。

3.1.1 医薬品の自発報告システムの有用性と薬剤疫学

自発報告システムは、投与された薬物とその適応症、発現した有害反応だけでなく、患者の性別、年齢、体重などの生理学的情報を収集しており、薬物による有害反応を包括的に解析する上で非常に有用なシステムである。米国食品医薬品局が管理している The FDA Adverse Event Reporting System (FAERS) は、2014年現在までに5百万件以上の薬物有害事象報告を収集しインターネット上で公開している。患者の幅広い集団にわたって薬物の使用や効果を監視したデータは、医薬品の有効性や安全性を研究する薬剤疫学（pharmacoepidemiology）に非常に貴重なリソースとなる。

これまでにFAERSのデータを解析した薬剤疫学分野の研究には、シグナル検出法（signal detection）と呼ばれる統計手法を駆使したものがいくつかある [7]。シグナルとは、添付文書には記載されていないような潜在的な副作用をいう。

例えば、Tamuraらは2004年から2009年までにFAERSに登録された1,644,220件

の薬物有害事象報告から、シグナル検出法で良く使われている proportional reporting ratio (PRR)、relative ratio (RR)、Information content (IC) および、Dumouchel により開発された gamma Poisson shrinker (GPS)[16] を用いてシグナル検出を行った。その結果、ワルファリン (Warfarin)、アスピリン (Aspirin)、クロピドグレル (Clopidogrel) などの抗血液凝固剤で、出血 (haemorrhage) や血腫 (haematoma) などの副作用のシグナルを検出した [60]。しかしながらこの研究では、解析対象とする薬物を抗血液凝固剤に限定し、適応症^{*11}の大きく異なる薬物にまたがる副作用を網羅的に解析するには至っていない。

次の3つの方法は、FAERSに加えて独立な情報を利用してシグナル検出手法の精度を高める工夫をしている。Harpazらは、FAERSのデータを電子健康記録 (electronic health record; EHR) と併用して使うことでシグナル検出の精度を高めた [27]。Vilarらは、2.1.1で述べたように、「化学構造の類似した薬物は類似した薬理作用を持つ」という見解に立ち、化学情報ベースのアプローチをシグナル検出法に組み込んでシグナルの検出力を高めた [64]。特に、横紋筋融解症に焦点を当て、FAERSからシグナル検出の方法のひとつであるGPSを用いて横紋筋融解症と強い関連を持つ薬物を検出した。スティーブンス・ジョンソン症候群 (Steven-Johnson syndrome; SJS) と中毒性表皮壊死症 (toxic epidermal necrolysis; TEN)^{*12} は、同じ発症機序を持ち、その多くが特定の薬物への免疫反応によって引き起こされると言われている。Papayらは、シグナル検出法のひとつである Multi-item gamma Poisson shrinker (MGPS) を用いてSJSとTENと関連のある5個の薬物を検出し、独立データである European case-control surveillance of severe cutaneous adverse reactions (EuroSCAR) と比較した [51]。しかしながら、これらの研究では特定の副作用に対する被疑薬を検出することに重点が置かれ、副作用発現の機序には言及していない。

3.1.2 本研究に至った経緯

添付文書に記載された副作用情報は、主に前臨床試験と臨床試験で得られたデータを元にしており、前臨床試験では、培養組織や実験動物で観察された毒性の情報が得られる。臨床試験では、実際の患者集団で観察された副作用の情報が得られるが、臨床試験に使われる被験者の集団は通常、妊婦、胎児、小児、高齢者などを含んでおらず、その規模も小

^{*11} 適応症 (indication for use): 薬物の使用目的。

^{*12} 中毒性表皮壊死症 (toxic epidermal necrolysis; TEN): 粘膜のびらんや広範囲にわたる皮膚の剥離を生じる重篤な皮膚病で、致死率はスティーブンス・ジョンソン症候群より高い。薬物による有害反応として起こる。

さい。そのため、まれな副作用や、個人間で発現の異なる副作用などは検出されにくいという問題がある。さらに、適応症の異なる薬物間での副作用プロファイルの比較はされない。一方、市販後調査では、製薬会社や医療機関により、薬物と副作用の情報だけでなく、患者の生理学的背景に関する情報も報告されるため、患者の個人差を視野に入れた解析が可能となる。本研究では、自発報告システムに登録された大規模な市販後薬物有害事象報告から、データマイニングの手法を用いて、開発段階では検出が難しいような副作用の情報の検出、特に患者の生理学的背景によって発症が異なるものを検出し、その機序を推定することを目的とした。

今までの自発報告システムデータの解析では、未知の薬物 副作用ペアの検出を目的として、シグナル検出法を用いた解析が行われてきた。しかしながら、ほとんどの研究では、解析対象を特定の薬物や副作用に限定し、異なる薬物間や副作用間での比較は行われてこなかった。さらに、副作用の発現機序の推定まで踏み込み、患者の背景による副作用のリスクなどは検討されてこなかった。

近年、生物学により得られた知見を収集提供する医薬品関連データベースが発達により、薬物の作用を生物学的な視点から捉えることが可能となった。本研究では、KEGG DRUG データベースから参照可能な知識データを利用することにより、副作用のクラスタのより詳細な特徴付けを行った。それでも不十分なときは、情報参照システムにより外部データベースを参照して、追加の情報を取得した。副作用に関しては、副作用を記述するために広く使われている医学用語の階層分類を用いて、検出された副作用を評価した。

このようなアプローチを取ることで、市販後薬物有害事象報告から検出された情報を生物学的な知識を用いて薬物の作用を分子レベルで詳細に特徴付けることで、分子レベルでの副作用発現機序の推定を目指した。さらに、患者のバックグラウンドにより異なる副作用を検出し、その因果関係を探ることを目的とした。

3.2 方法

3.2.1 市販後有害事象報告に登録された薬物と副作用のデータ

FAERS には、2004 年第 1 四半期から 2012 年第 3 四半期の 35 期間に約 4 百万件の薬物有害事象報告が登録されている。薬物有害事象報告は、表 10 のように 7 種類のデータから構成され、データの種類ごとに 7 つのファイルに分けて配布されている。本研究では、そのうちの *DEMO*、*DRUG*、*REAC*、*INDI* の 4 種類のファイルに登録されたデータを使用した。

表 10 FAERS の 7 つのファイル構成

ファイル名	データの内容	報告件数	本研究で使用
<i>DEMO</i>	人口統計学的・行政的データ	4,279,732	
<i>DRUG</i>	薬物に関するデータ	4,280,321	
<i>REAC</i>	有害反応（副作用）データ	4,280,312	
<i>OUTC</i>	患者の予後データ	3,182,555	
<i>RPSR</i>	報告のデータソース	971,935	
<i>THER</i>	治療に関するデータ	2,887,413	
<i>INDI</i>	薬物の適応症のデータ（治療目的）	3,464,034	

DRUG ファイルは、投与された薬物についてのデータを登録している。該当する報告件数は 4,280,321 件である。薬物は、第一被疑薬（primary suspect drug; PS）、第二被疑薬（secondary suspect drug; SS）、併用薬（concomitant drug; C）などに分類されており、本研究では、PS 又は SS に分類された薬物のうち、他の被疑薬との併用投与が無く、単体で投与されたもののみを用いた。その結果、4,939 個の薬物名を含む 1,670,741 件の報告が該当した。

DRUG ファイルを扱う際には、薬品名を用いて個々の薬物を識別する必要があるが、薬品名は登録者が入力した通りに記載されているものもあるので、そのような薬物はデータから除き、FDA によって表記が標準化されている薬物のみを使用した。計算処理を可能にするため、薬物は、KEGG DRUG データベースで薬物の識別番号として使われている D 番号に統一した。統一は、FAERS の薬品名と KEGG DRUG の薬物名とを正規表現で完全マッチさせることにより行った。この作業により、FAERS の 1,909 個の薬品名が KEGG DRUG の 1,384 個の D 番号に変換され、その結果 483,071 件の報告が残っ

た。この手順により、KEGG DRUG 内に蓄積されている様々な医薬品関連情報を用いて薬物の注釈付けを行うことが可能となる。中でも、医薬品分類 Anatomical Therapeutic Chemical (ATC) classification system による薬物の階層的分類や、標的タンパク質の生物学的な情報を使用できるようになる。さらに、KEGG の情報参照システムである LinkDB[21] を利用し、DrugBank, SIDER, DailyMed などの外部の医薬品関連データベースを相互参照することが可能となる。

REAC ファイルは、薬物の有害反応(副作用)のデータを登録している。該当する報告数は 4,280,312 件である。INDI ファイルは、薬物の適応症のデータ(治療目的)を登録している。該当する報告数は 3,464,034 件である。INDI ファイルと REAC ファイルの両方で同じ症状を表す用語が含まれている報告はデータから除いた。

REAC ファイルの副作用キーワードと INDI ファイルの適応症キーワードは、後述の Medical Dictionary for Regulatory Activities (MedDRA) の “Preferred Terms” (PT) 階層の用語で記述されている(表 11)。副作用キーワードは、PT 階層の用語では後のバイクラスタリング解析には細かすぎるため、ひとつ上の階層である “High Level Terms” (HLT) 階層の用語にまとめた。さらに、“System Organ Class” (SOC) 階層で *Investigations*(調査)、*Social circumstances*(社会状況)、*Surgical and medical procedures*(外科および内科処置)に分類される副作用キーワードはデータから除いた。これらの用語は FAERS に副作用キーワードや適応症キーワードとして登録されていたとしても、薬物との因果関係を解釈するのが難しいためである。適応症キーワードに対しても同様の作業を行った。

その結果、1,317 個の副作用キーワード (MedDRA HLT 階層の用語) および 1,374 個の薬物 (KEGG D 番号) からなる 692,707 件の有害事象 (報告数は 425,273 件) が残った。このデータから、行を副作用、列を薬物とする有害事象の頻度行列を作成した。

表 11 MedDRA の階層分類と用語数

階層	英語表記	日本語表記	用語数
SOC	System Organ Class	器官別大分類	26
HLGT	High Level Group Terms	高位グループ語	335
HLT	High Level Terms	高位語	1,713
PT	Preferred Terms	基本語	19,737
LLT	Lowest Level Terms	下層語	70,634

MedDRA は日米欧医薬品規制ハーモナイゼーション国際会議 (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; ICH) で開発された国際医薬用語集であり、副作用表記に広く利用されている (<http://www.meddra.org/>)。表 11 にあるように、薬物の副作用や適応症を表す用語を階層的に 5 段階に分類している。本研究では MedDRA バージョン 15.1 を使用した。

3.2.2 バイクラスタリング

バイクラスタリングはデータマイニングの分野で近年盛んに研究され、複数のアルゴリズムが開発されている。バイオインフォマティクスの分野ではマイクロアレイ遺伝子発現解析などに使用されている [32]。バイクラスタリングは、データ行列から、類似した数値変動パターンを持つ部分行列を抽出する。ここでは、行を副作用、列を薬物とする有害事象の頻度行列から、副作用の部分集合と、該当する薬物の部分集合を同定する。この二つの部分集合をあわせてバイクラスタと呼ぶ。バイクラスタリングのためのアルゴリズムは多数あり、それぞれ一長一短があるが、本研究では Iterative Signature Algorithm (ISA) [31] を使った。このアルゴリズムの特徴は、サイズが大きくノイズも多い行列に対しても適用可能であることであり [11, 31]、前述のマイクロアレイ遺伝子発現解析にも用いられていることから、本研究で扱ったデータ行列にも適用可能であると判断した。しかしながら、このアルゴリズムでは出力される複数のバイクラスタ間に重なりを許すため、パラメータの最適化などには注意を要する。

3.2.3 Iterative Signature Algorithm (ISA)

本研究で用いた ISA について説明する。行（ここでは副作用）の集合 R_{all} と列（ここでは薬物）の集合 C_{all} からなる行列に対し、行の部分集合と、関連する列の部分集合から構成される部分行列をバイクラスタという。ISA では、ランダムに選択した行の部分集合と列の部分集合から出発して、以下のプロセスを反復的に行い、最適なバイクラスタを求める（図 13）。バイクラスタの厳密性を調整するパラメータ $thr.row$ および $thr.col$ はユーザーが与える。

- Process 1. ランダムに $r \in R_{all}$ をいくつか選択し、この集合を $R^{(0)}$ とおく。
- Process 2. 列に関するスコア $s_c = \langle s_r E_r \rangle_{r \in R^{(0)}}$ を各列について計算し、 $s_c > thr.col$ を満たす列の集合 $C^{(0)} \subseteq C_{all}$ を得る。ここで $\langle s_r E_r \rangle$ は、列をひとつ固定したときに $R^{(0)}$ に属する全ての r でその値の加重平均を取ったものである。
- Process 3. 行に関するスコア $s_r = \langle s_c E_c \rangle_{c \in C^{(0)}}$ を各行について計算し、 $s_r > thr.row$ を満たす行の部分集合 $R^{(1)} \subseteq R_{all}$ を得る。ここで $\langle s_c E_c \rangle$ は、行をひとつ固定したときに $C^{(0)}$ に属する全ての c でその値の加重平均を取ったものである。
- Process 4. Process 2, Process 3 を反復的に行い、最終的に $R^{(n+1)} = R^{(n)}$ に収斂した時点で反復を止め $R^{(n)}, C^{(n)}$ を出力する。

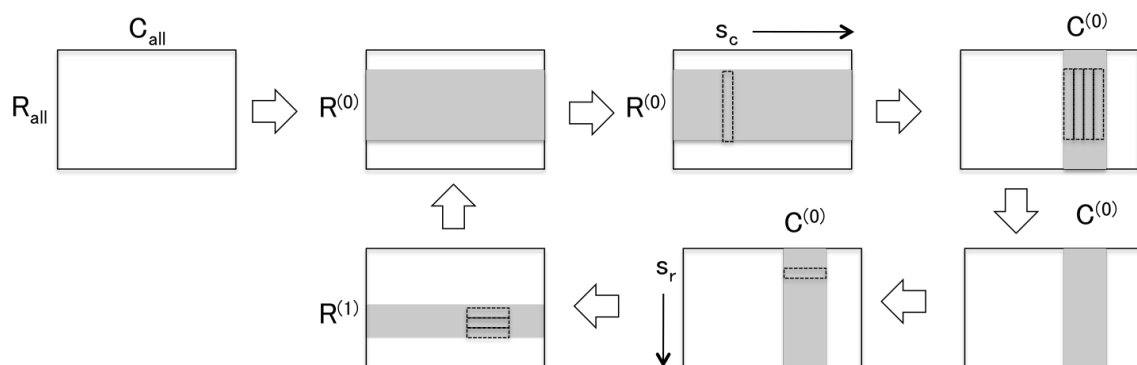


図 13 Iterative Signature Algorithm (ISA) の手順

3.2.4 バイクラスタの同定

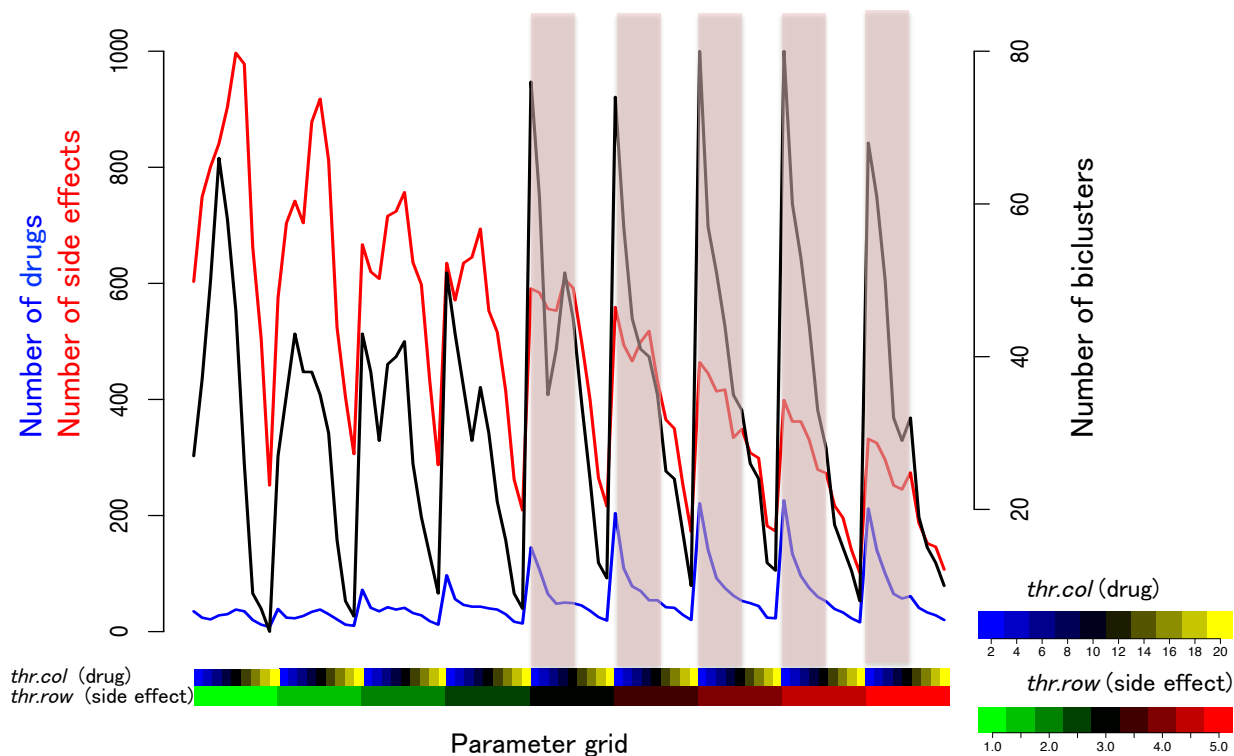
バイクラスタリングでは一般に、バイクラスタの厳密性とカバレッジとの間にトレードオフを伴う。すなわち、バイクラスタの厳密性を上げようとするれば、そのサイズは小さくなり、データ全体に占める割合も低くなる。逆に厳密性を下げればカバレッジは上がるが、バイクラスタには多くのノイズが含まれることになる。ここで言う厳密性とは、バイクラスタ内の数値変動をどれくらい均一にするかの程度をいう。さらには、バイクラスタリングは教師無し的手法であるため、厳密性を調整するパラメータを最適化する一般的な枠組みが無い。そのため、データから意味のある情報を抽出するには、バイクラスタの厳密性とカバレッジとのバランスをモニタリングして慎重に行う必要がある。本研究では [32] の手法を採用し、ISA のワークフロー内で使用するパラメータ $thr.row$ 、 $thr.col$ を複数設定することにした。

まず、3.2.1 で作成した、副作用を行、薬物を列とする $1,374 \times 1,317$ 行列を行ごとに正規化した。ここでは、異なる副作用間で数値変動のパターンを比較するために、行ごとに正規化を行った。次に、行の閾値である $thr.row$ を 1 から 5 の範囲で 0.5 刻みに設定した。列の正規化はせず、そのため行より閾値の幅を広く取り、 $thr.col$ は 2 から 20 の範囲で 2 刻みに設定した。また、開発者の奨励するフィルタリング手法を用いて堅牢なもの（再現性の高いもの）をフィルタリングした。

$thr.row$ と $thr.col$ の最適値は、バイクラスタのカバレッジと厳密性をモニタリングしながら慎重に決定した。ここでは、全バイクラスタに抽出される副作用の総数と薬物の総数をカバレッジの指標とし $thr.row$ と $thr.col$ を上げ、最終的に行側の閾値を $3 \leq thr.row \leq 5$ に、列側の閾値を $2 \leq thr.col \leq 10$ に決定した（図 14）。

異なる複数の閾値を用いてバイクラスタを抽出したため、得られたクラスタは非常に冗長であった。そのため、ISA アルゴリズムの冗長除去関数を用いて冗長なバイクラスタを削除した（図 15）。このとき、類似度が 0.8 であるバイクラスタ同士を同一のものとした。

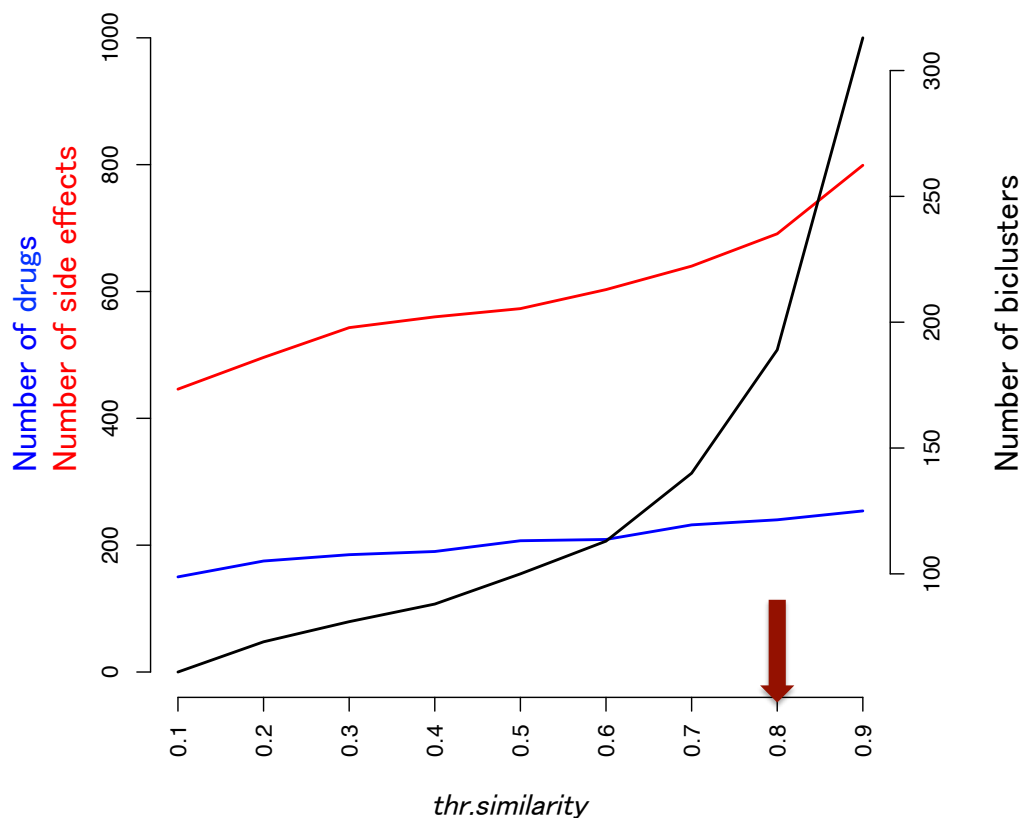
Number of biclusters, drugs and side effects at different sets of *thr.col* & *thr.row*



Optimal parameters were set to $3 \leq thr.row \leq 5$ and $2 \leq thr.col \leq 10$

図 14 ISA アルゴリズムでのパラメータ最適化 1
 同定されるバイクラスタの厳密性を調整する 2 つのパラメータのうち、行のパラメータ *thr.row* を 1 から 5 の範囲で 0.5 刻みに設定し、列のパラメータ *thr.col* を 2 から 20 の範囲で 2 刻みに設定した。*thr.row* と *thr.col* が取りうる全 90 通りの数値の組み合わせのそれぞれに対して、同定されたバイクラスタの数 (黒) をプロットした。バイクラスタ数は各 *thr.row* の中で *thr.col* の増加に伴って増減し、その増減パターンは *thr.row* が 3 以上で比較的安定する。また、該当するバイクラスタ数に現れる副作用の総数 (赤) と薬物の総数 (青) をプロットしたところ、*thr.row* が増加するに従って副作用数が減少する傾向は明らかであるが、それぞれの増減の中で特に *thr.col* = 10 を超える付近で副作用数が急速に減少する。このことから、副作用の総数を保持しながらバイクラスタの厳密性を高めるためのパラメータの最適値として、 $3 \leq thr.row \leq 5$ 、 $2 \leq thr.col \leq 10$ と定めた。

Number of biclusters, drugs and side effects at different sets of *thr.similarity*



Optimal parameter was set to *thr.similarity* = 0.8

図 15 ISA アルゴリズムでのパラメータ最適化 2

バイクラスタの厳密性を調整する 2 つのパラメータである *thr.row* と *thr.col* を複数の値で設定したために、冗長なバイクラスタが同定された。冗長性の尺度として、バイクラスタ間の類似性スコア *thr.similarity* を定め、*thr.similarity* を 0.1 から 0.9 の範囲で調整した。そのときのバイクラスタの数 (黒) と、該当する副作用の総数 (赤色) および薬物の総数 (青) をプロットした。副作用の総数と薬物の総数をできるだけ多く保持しながら冗長バイクラスタを削除する必要があるため、バイクラスタの数が急速に増加する *thr.similarity* = 0.8 を最適な類似度スコアと定めた。

3.2.5 副作用キーワードの医学的分類体系を用いた副作用クラスタの評価

3.2.1 で記したように、本研究では、副作用を MedDRA の HLT 階層（表 11）の用語で記述した。バイクラスタリングにより各副作用クラスタに抽出された副作用キーワードの類似性を検証するために、各副作用キーワードを、HLT 階層のひとつ上の階層である High Level Group Terms (HLGT) 階層にマップした。その上で、マップの不均一性をジニ係数を使用して測定し、クラスタ内の副作用キーワードの類似性の指標とした。ジニ係数は集団内の不均一性を測る指標で、社会学の分野で所得分布の不平等さを測る指標として使われている。ジニ係数は 0 から 1 までの値をとり、集団の分布が均一であれば 0 に近づき、不均一であれば 1 に近づく。すなわち、高いジニ係数は特定のカテゴリへの分布の偏りを示す。

3.2.6 薬物の適応症を用いた副作用クラスタの評価

副作用クラスタに抽出された薬物の多様性を検証するために、FAERS の *INDI* ファイル（表 10）から薬物の適応症データを取得した。163 個のクラスタに含まれた 182,000 件の報告のうち、95,888 件の報告に関して適応症のデータが得られた。適応症データは MedDRA の PT 階層で記述されているので、MedDRA の HLT 用語に変換した。クラスタ内の薬物について、適応症ごとに報告件数をカウントした適応症プロファイルを作成した。その上で、適応症プロファイルのペアワイズなピアソン相関係数を計算し、クラスタ内の全てのペアにわたる平均を取り、クラスタ内の薬物の多様性の指標とした。163 個のクラスタの薬物の適応症プロファイルは補足表 S4 (<http://web.kuicr.kyoto-u.ac.jp/supp/smizutan/adr-indication/TableS1.txt>) に示す。

3.2.7 患者の性別・年齢・体重の情報を用いた副作用クラスタの比較

クラスタ間で共有する副作用に患者の生理学的な背景が寄与している可能性を検証するために、FAERS の *DEMO* ファイル（表 10）から患者の性別、年齢、体重の情報を取得した。163 個のクラスタに関連した合計 182,000 有害事象報告のうち、性別に関するデータが得られた報告は 166,704 件、年齢に関するデータが得られた報告は 80,013 件、体重のデータが得られた報告は 33,315 件であった。次に、異なる副作用を持つクラスタのうち、共通の薬物を持つクラスタ同士を比較し、それぞれのクラスタに該当する患者の生理学的な背景に違いがあるかを検証した。性別の分布の違いはフィッシャーの正確確率検定により統計的有意性を検証し、年齢および体重の分布はウェルチの *t* 検定により統計的有意性を検証した。

3.3 結果と考察

3.3.1 バイクラスタリングによる副作用クラスタの同定

副作用を行、薬物を列とする $1,374 \times 1,317$ 有害事象頻度行列を作成し、有害事象の発現頻度の数値変動パターンにより副作用をクラスタリングするために、ISA を用いてバイクラスタリングを行った。その結果、691 の副作用と 240 の薬物を含む合計 163 個のバイクラスタが同定された。これらのバイクラスタには、182,000 の報告 (データ行列の 425,273 報告の 43%) が含まれていた。本研究では、以降、これらを副作用クラスタと呼ぶ。ISA の性質と、複数のパラメータ設定により、一部の副作用と薬物は複数のクラスタに現れていた。副作用クラスタには、類似した有害事象頻度パターンを持つ副作用と、それに関連する薬物が抽出されていることが期待される。図 16 は 163 個のバイクラスタの各々に現れる副作用数と薬物数を示している。副作用と薬物の詳細情報は補足表 S4 (<http://web.kuicr.kyoto-u.ac.jp/supp/smizutan/adr-indication/TableS1.txt>) に示す。

3.3.2 バイクラスタリング後の副作用と薬物の種類の評価

副作用クラスタに抽出された副作用および薬物は、ある特定の種類の偏っているかもしれない。その場合には、副作用や薬物の多様性を解析することは期待できないであろう。そこで、データ行列の多様性が副作用クラスタに保存されているかを、MedDRA の分類体系と ATC の分類体系を用いて評価した。副作用と薬物の総数を、バイクラスタリング前後で比較した結果、163 個のクラスタに現れる副作用と薬物の総数は少なくなっていたものの、分類的な多様性は保持されていた (図 17)。このことは、バイクラスタリングの手順によっても副作用や薬物の種類の多様性が失われていないことを示す。

Number of drugs and side effects extracted in 163 biclusters

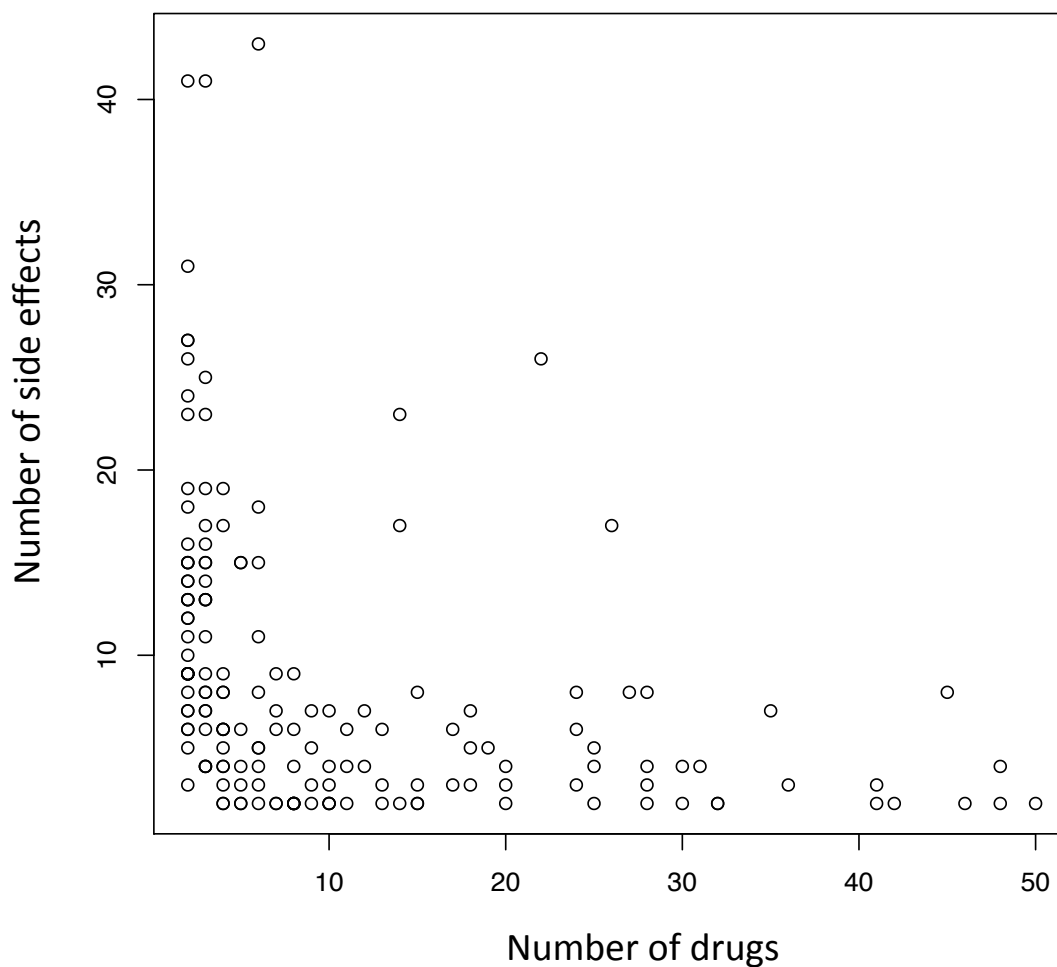


図 16 163 個の副作用クラスターの薬物と副作用の数 (Mizutani ら [44] より引用)
各点は 163 個の副作用クラスターを表し、各クラスターが持つ薬物の数を横軸に、副作用の数を縦軸にプロットした。163 個のクラスターの副作用と薬物に関する詳細な情報は補足表 S4 (<http://web.kuicr.kyoto-u.ac.jp/supp/smizutan/adr-indication/TableS1.txt>) に示す。

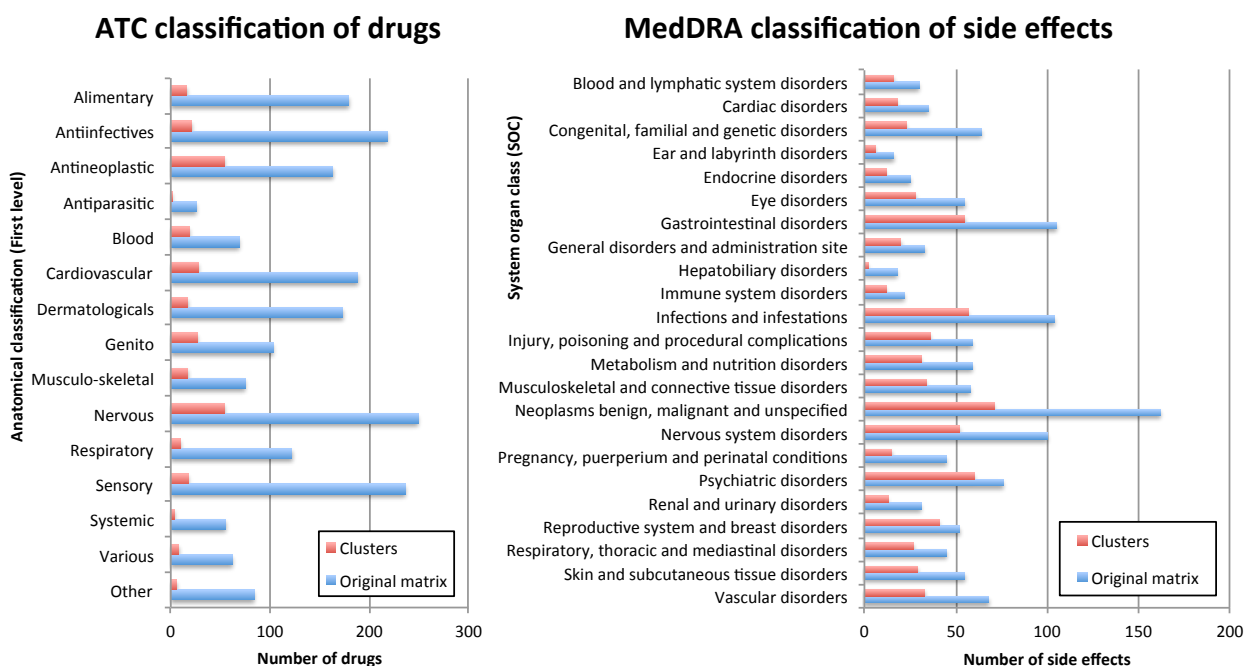


図 17 副作用クラスタにおける薬物の ATC 分類と副作用の MedDRA 分類 (Mizutani ら [44] より引用)

(左) 薬物の数を ATC 解剖学的レベル (第一階層) のカテゴリごとに集計した。

(右) 副作用の数を MedDRA の SOC レベル (第一階層) のカテゴリごとに集計した。

1,374 × 1,317 データ行列 (青) と 163 個の副作用クラスタ (赤) で比較すると、バイクラスタリング後では薬や副作用の総数は減少しているが、ATC 分類と MedDRA 分類での多様性は保存されている。

3.3.3 添付文書への参照による副作用クラスタの評価

163 個の副作用クラスタに抽出された情報が添付文書にどの程度記載されているかを評価するために、SIDER データベースを参照した。元のデータ行列では、少なくとも一件の報告を持つ 80,019 の薬物 副作用対のうち、34,315 対 (43%) が SIDER データベースに登録されていた。それに対し、バイクラスタリング後では、163 個のクラスタの少なくともひとつに現れる 5,159 の薬物 副作用対のうち 2,481 対 (48%) が SIDER データベースに参照された。このことは、バイクラスタリングによってデータ内に存在する既存の知識の割合が失われていないことを指摘している。

3.3.4 副作用クラスタ内の副作用と薬物の特徴付け

データ行列を行ごと（副作用ごと）に見ると、その行の値の和はこの副作用を発症した患者の総数であり、各値は、それぞれの薬物の投与件数を表している。副作用クラスタには、バイクラスタリングにより、副作用の発現頻度パターンが類似して多い薬物同士が同じクラスタに集められている。また、副作用クラスタに抽出された副作用は全て、該当する薬物について同様に高い発現頻度パターンを持つ。すなわち、副作用クラスタの薬物は、関連する副作用の発現に関して何らかの共通性を持っているのではないかと考えられる。そこで、各副作用クラスタに抽出された副作用の集合と、薬物の集合について以下のような検証を行った。

まず、クラスタ内の副作用が互いに医学的類似性を持っているかを MedDRA の分類体系を用いて調べた（図 18）。図 18（右）のヒストグラムは 163 個の副作用クラスタの副作用のジニ係数の分布であり、そのうちの大部分は比較的高いジニ係数を示しているのが分かる。ジニ係数は集団内の不均一性を測る指標で 0 から 1 の間を取り、1 に近ければ近いほどその集団が特定のカテゴリに偏りを持つことを示す。すなわち、副作用クラスタの大部分は医学的に類似した副作用を持つことを示唆している。

次に、163 個のクラスタに抽出された薬物の類似性を、その適応症の視点から検証した。自発報告システムでは、薬物投与が原因で発症した有害反応データの中に患者の元の症状（すなわち薬物の適応症）が含まれる場合がある。そのため、クラスタに抽出された副作用は、該当する薬物の共通した適応症である場合が考えられる。この可能性を排除するために、クラスタに抽出された薬物の適応症の類似性を検証した。薬物の適応症の多様性が示されれば、副作用は薬物が引き起こした真の有害反応である可能性が高いと示唆される。薬物の適応症の類似度は、FAERS に提供されている適応症情報を元にピアソン相関係数を用いて算出した。図 18（上）のヒストグラムは、163 個のクラスタのうち 145 個（89%）でピアソン相関係数が 0.5 未満であったことを示している。すなわち、クラスタ内の薬物は互いに異なる治療目的のために投与されたことが分かる。

また、図 18（中央）の散布図は、145 個のクラスタが残りの 18 個のクラスタから分離されることを示している。これらの結果から、バイクラスタリングにより同定された 163 個のクラスタの内の 145 個は、医学的類似性の高い副作用が異なる適応症を有する薬物の投与の結果として発症する傾向を示すカタログとして見ることができる。

Intra-cluster characterization of drugs and side effects

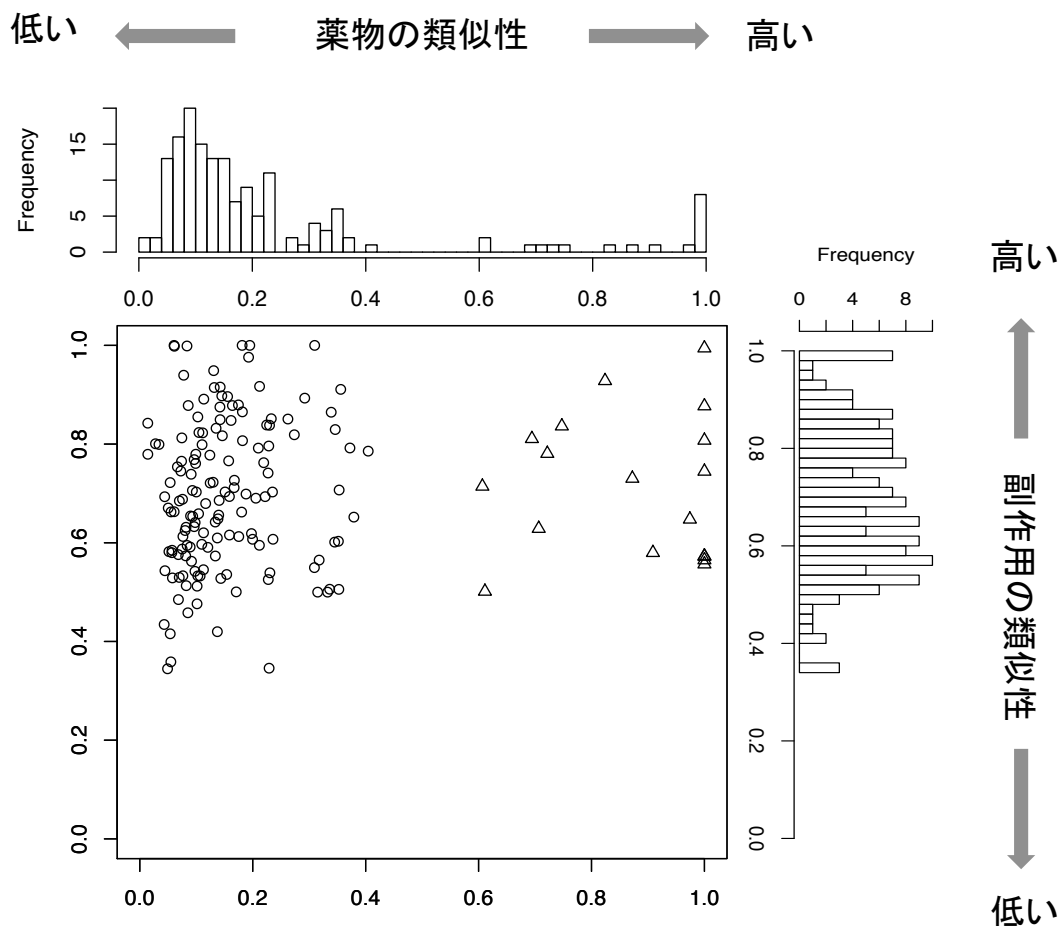


図 18 163 個の副作用クラスターの薬物と副作用の特徴付け (Mizutani ら [44] より改変)

163 個の副作用クラスターの各々について、副作用の医学的類似性と薬物の適応症の多様性を定量化した。(上) 薬物の適応症プロファイルに基づくピアソンの相関係数の分布。163 個のクラスターのうち 145 個 (89%) はピアソンの相関係数が 0.5 未満であった。(右) 副作用の医学的類似性の定量化に用いたジニ係数の分布。副作用クラスターの大部分は高いジニ係数を示した。(中) ピアソンの相関係数 (横軸) とジニ係数 (縦軸) に対して 163 個の副作用クラスターをプロットした。ピアソンの相関係数が 0.5 未満であった 145 個のクラスター (丸) は残りの 18 個のクラスター (三角) から分離された。これらの 145 個の副作用クラスターは、多様な適応症プロファイルを持つにも関わらず、医学的類似性の高い副作用を持つことを示唆している。

3.3.5 副作用クラスタの例

145 個のクラスタからは、異なる適応症を持つ薬物が医学的に類似した副作用を発現する機構を推定する材料となることが示唆される。その例として 3 つのクラスタについて議論する (表 12 および 図 19)。

第 17 クラスタには、*Gastrointestinal haemorrhages* や *Nervous system haemorrhagic disorders* など、主に消化管や神経系の出血性障害の副作用キーワードが抽出された。このクラスタでこれらの副作用キーワードと関連する薬物は、おおまかに分けて次の 3 つのタイプの適応症を持っていることが分かった。適応症情報は、このクラスタに該当する有害事象報告のデータ (INDI ファイル) から得た。

Type 1

ワルファリン (Warfarin) は抗血液凝固薬 (ビタミン K エポキシド還元酵素阻害剤) である。このクラスタに関連する有害反応報告では主に *Supraventricular arrhythmias* (上室性不整脈) の治療に用いられていた。

Type 2

ロフェコキシブ (Rofecoxib) と セレコキシブ (Celecoxib) は非ステロイド性抗炎症薬 (NSAID) (選択的 COX-2 阻害剤) であり、*Arthropathies* (関節症)、*Osteoarthropathies* (骨関節症) およびその他の筋骨格や結合組織の鎮痛剤として使われていた。

Type 3

アスピリン (Aspirin) とクロピドグレル (Clopidogrel) は、*Coronary necrosis and vascular insufficiency* (中枢神経系の出血や脳血管障害) や *Cerebrovascular and spinal necrosis and vascular insufficiency* (冠動脈壊死および血管不全) の治療に使われていたが、これらの薬は別の効能を持っている。アスピリンは抗血栓剤であるが、NSAID (非選択的 COX-1,2 阻害剤) でもある。クロピドグレルもまた抗血液凝固薬であるが、アスピリンとは異なる受容体 (P2Y₁₂ 受容体) に対する拮抗薬である。

ワルファリンによる出血のリスクは、投与量の調整が不十分であるために起こるとされている [1, 40]。また、出血のリスクはアスピリンやクロピドグレルを含むいくつかの NSAID についても報告されている [13, 60]。実際、アスピリン、クロピドグレル、ロフェコキシブなどの NSAID は、COX の活性を阻害することにより血小板の血液凝固因子であるトロンボキサン A₂ (thromboxane A₂) の産生を抑制することが推定される。すな

わち、第 17 クラスタに抽出された消化管や神経系の出血障害は、図 19A に挙げたように、抗血液凝固薬の不完全な調整により引き起こされた副作用であると解釈できる。

第 29 クラスタに抽出された副作用キーワードは *Partial vision loss* (部分視覚喪失)、*Retinal structural change, deposit and degeneration* (網膜構造変化、沈着および変性) および *Ocular haemorrhagic disorders* (眼部出血性障害) など、全て眼障害を表すものだった。このクラスタに関連する薬物は次に挙げるように多様な適応症を有していた。

Type 1

ベバシズマブ (Bevacizumab) およびラニビズマブ (Ranibizumab) は、血管内皮増殖因子 (vascular endothelial growth factor; VEGF) の活性を阻害するモノクローナル抗体 IgG1 である。このクラスタでは、*Retinal structural change, deposit and degeneration* (網膜構造変化、沈着および変性)、*Ocular vascular disorders* (眼部血管障害) および *Choroid and vitreous haemorrhages and vascular disorders* (脈絡膜および硝子体出血および血管障害) の治療に使用されていた。ベルテポルフィン (Verteporfin) は上記の 2 つの生物学的製剤とは異なり、小分子薬であるが、この薬物も上記の 3 つの適応症を持っていた。

Type 2

血管作動薬であるシルデナフィル (Sildenafil) は、このクラスタでは *Sexual arousal disorders* (性衝動障害) の治療に使われていた。

Type 3

インターフェロン β -1a (Interferon beta-1a) は、*Multiple sclerosis acute and progressive* (急性および進行性多発性硬化症) の治療に使われていた。

Type 4

エタネルセプト (Etanercept) は、誘発性サイトカインである腫瘍壊死因子 (tumor necrosis factor; TNF) に作用し、TNF を循環器系から除去することによって炎症を軽減する生物学的製剤であり、このクラスタでは *Rheumatoid arthritis and associated conditions* (関節リウマチおよびその関連疾患) および *Psoriatic conditions* (乾癬状態) の緩和に使用されていた。

このクラスタでは、MedDRA の HLT レベルでの副作用と適応症を表す用語に共通するものが含まれていた。ベルテポルフィン、ベバシズマブ、ラニビズマブは眼組織の構造変化と出血治療に使用されていたが、この適応症を記述する用語は副作用の用語としても報告されていた。このことは、これらの薬物治療が十分な有効性を持たなかったか、または元の症状を悪化させた結果、ここに挙げたような眼関連の症状の発現につながったことを

示唆している。特に、ベバシズマブとラニビズマブが眼に関わる有害反応を引き起こすことは既に指摘されている [63]。このクラスタに関するもうひとつの注意点は、血管の形成および調節における予想外の制御の関与を示唆していることである。シルデナフィルの標的タンパク質である cGMP ホスホジエステラーゼ (cGMP phosphodiesterase) は網膜にも発現することが知られている。これらは、異なる薬物によって引き起こされる眼障害の発生が共通のメカニズムに起因していることを示唆している (図 19B)。

第 30 クラスタでは、運動に関する副作用キーワードが抽出され、関連する薬物の適応症は 2 種類に大別される。

Type 1

アリピプラゾール (Aripiprazole)、ハロペリドール (Haloperidol)、クエチアピン (Quetiapine)、メシル酸ジプラシドン (Ziprasidone mesylate)、塩酸ジプラシドン水和物 (Ziprasidone hydrochloride) および、リスペリドン (Risperidone) は、5-ヒドロキシトリプタミン (セロトニン) 受容体 (5-hydroxytryptamine (serotonin) receptor; 5-HT) および/またはドパミン受容体 D2 の拮抗薬である。これらは全て類似した適応症のプロファイルを示した。ジプラシドンとリスペリドンは、*Schizophrenia* (統合失調症) および *Bipolar disorders* (双極性障害) の治療に使われる非定型抗精神病薬である。アリピプラゾールとハロペリドールもこれらの障害の治療に用いられていたが、一般には、他の精神疾患にも適用される。クエチアピンは、*Schizophrenia* (統合失調症) の治療に使用されていた。この薬物は多元受容体標的抗精神病薬 (multi-acting receptor targeted antipsychotics; MARTA) と言われ、セロトニン受容体やドパミン受容体の他にも多くの受容体に作用する。

Type 2

アトモキセチン (Atomoxetine) は上記の薬とは異なる適応症プロファイルを示した。この薬物は選択的ノルエピネフリン再取り込み阻害剤 (norepinephrine reuptake Inhibitors; NRI) であり、*Attention deficit and disruptive behaviour disorders* (注意欠陥多動性障害) の治療に使用されていた。

これらの薬物は全て、運動に関わる副作用を発現しており、具体的には *Dyskinesias* (ジスキネジア^{*13})、*Dystonias* (ジストニア^{*14})、*Altered body temperature* (体温変動)

^{*13} ジスキネジア (Dyskinesias): 不随意運動の一種で、自分の意志に関わりなく身体が動いてしまう症状をいう。

^{*14} ジストニア (Dystonias): 中枢神経系の障害による不随意で持続的な筋収縮にかかわる運動障害をいう。

Parkinsonism (パーキンソニズム^{*15})、*Eyelid movement disorders* (眼瞼運動障害) および、*Decreased physical activity levels* (身体の活動性レベルの減少) などの副作用が抽出された。このことから、このクラスタは、2 つタイプの薬物が運動異常の発生に関与していることを示唆している。ジスキネジアのような運動障害は、ドパミン作動性閉塞が原因であると疑われてきた [43]。タイプ 1 の薬物は、ドパミン受容体を阻害する。ノルエピネフリントランスポーターもまた前頭前野でのドパミンの取り込みに関与していることが報告されている [46]。総じて、このクラスタで抽出された運動障害の発現の機序には、ドパミン受容体の遮断に寄与していることが示唆される (図 19C)。

^{*15} パーキンソニズム (Parkinsonism): パーキンソン病の典型的な症状である振戦、筋固縮、無動、振戦、姿勢反射・保持障害のうちのいくつかを示している症状をいう。

表 12 副作用クラスタの例 (Mizutani ら [44] より改変)

Cluster ID	Side effects	Drugs	Indications for use
Cluster 17 ADR score = 0.6987666 drug score = 0.1155793	Gastrointestinal haemorrhages Non-site specific gastrointestinal haemorrhages Nervous system haemorrhagic disorders Cerebral injuries Traumatic central nervous system haemorrhages Gastric and oesophageal haemorrhages	Type 1: Warfarin (D00564)	Supraventricular arrhythmias Peripheral embolism and thrombosis Cardiac valve therapeutic procedures
		Type 2: Rofecoxib (D00568) Celecoxib (D00567)	Arthropathies Osteoarthropathies Musculoskeletal and connective tissue pain and discomfort
		Type 3: Aspirin (D00109) Clopidogrel (D07729)	Coronary necrosis and vascular insufficiency Cerebrovascular and spinal necrosis and vascular insufficiency Prophylactic procedures
Cluster 29 ADR score = 0.5255139 drug score = 0.1164479	Partial vision loss Retinal structural change, deposit and degeneration Ocular haemorrhagic disorders Eye injuries Choroid and vitreous haemorrhages and vascular disorders	Type 1: Bevacizumab (D06409) Ranibizumab (D05697) Verteporfin (D01162)	Retinal structural change, deposit and degeneration Ocular vascular disorders Choroid and vitreous haemorrhages and vascular disorders
		Type 2: Sildenafil (D02229)	Sexual arousal disorders General signs and symptoms
		Type 3: Interferon beta-1a (D04554)	Multiple sclerosis acute and progressive
		Type 4: Etanercept (D00742)	Rheumatoid arthritis and associated conditions Psoriatic conditions
Cluster 30 ADR score = 0.9108681 drug score = 0.3415874	Dyskinesias and movement disorders Dystonias Body temperature altered Parkinson ' s disease and parkinsonism Eyelid movement disorders Decreased physical activity levels	Type 1: Aripiprazole (D01164) Haloperidol (D00136) Quetiapine (D00458) Ziprasidone mesylate (D02100) Ziprasidone hydrochloride (D01939) Risperidone (D00426)	Schizophrenia Bipolar disorders Psychotic disorder General signs and symptoms Schizoaffective and schizophreniform disorders Depressive disorders
		Type 2: Atomoxetine (D02574)	Attention deficit and disruptive behaviour disorders

(列 1) ADR score は副作用の医学的類似性の指標として用いたジニ係数の値。drug score は薬物の適応症プロファイルのピアソン相関係数。(列 2) クラスタに抽出された副作用の一覧。(列 3) クラスタに抽出された薬物の一部を適応症別に分類した。() 内は KEGG DRUG の D 番号。(列 4) 薬物のタイプごとの適応症。

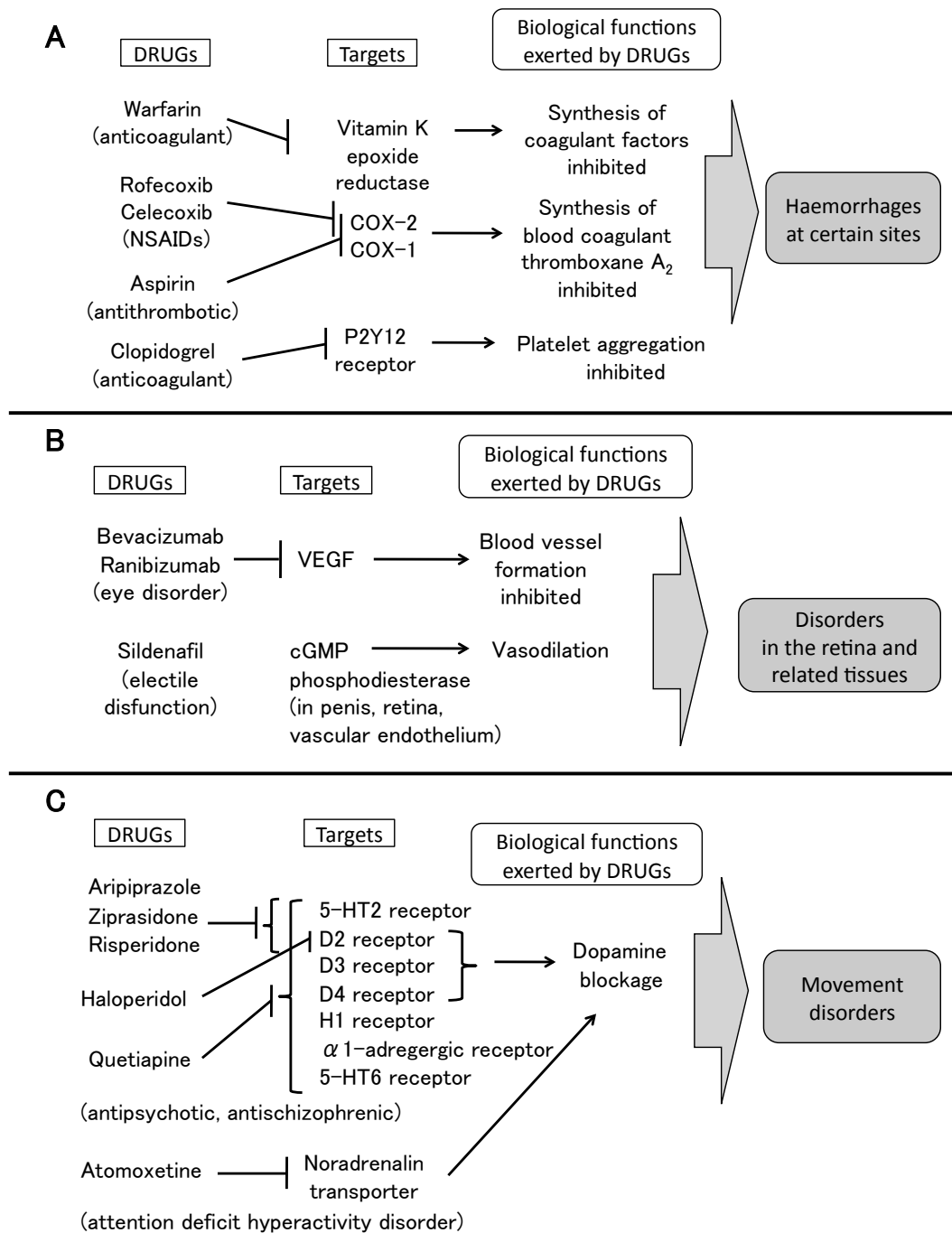


図 19 副作用クラスタからの副作用の発現機序推定 (Mizutani ら [44] より引用)
 (A) 第 17 クラスタの副作用 Haemorrhages at certain sites の発現機序
 (B) 第 29 クラスタの副作用 Disorders in the retina and related tissues の発現機序
 (C) 第 30 クラスタの副作用 Movement disorders の発現機序
 本文中で副作用の機序を詳細に推定した。副作用の詳細は表 12 に示す。

3.3.6 患者の生理学的背景に基づいた副作用クラスタの特徴付け

副作用クラスタが患者の生理学的背景で特徴付けられるケースについて調査した。そのようなケースを検出するために、次の手順で副作用クラスタのペアワイズな比較を行った。

バイクラスタリングにより同定された 163 個の副作用クラスタには、互いに共通な薬物を持つものがある。同様に、互いに共通な副作用を持つものもある。図 20 のように、同一の薬物に対して互いに排他的な副作用を有する副作用クラスタのペアに注目した。このような条件を満たすクラスタのペアは 17,363 対あり^{*16}、合計 152 個の薬物に関係していた。このようなクラスタのペアで、もし患者の生理学的背景に違いが見られれば、各クラスタに表れる副作用群の背景には患者の生理学的特徴が関わっている可能性が示唆される。

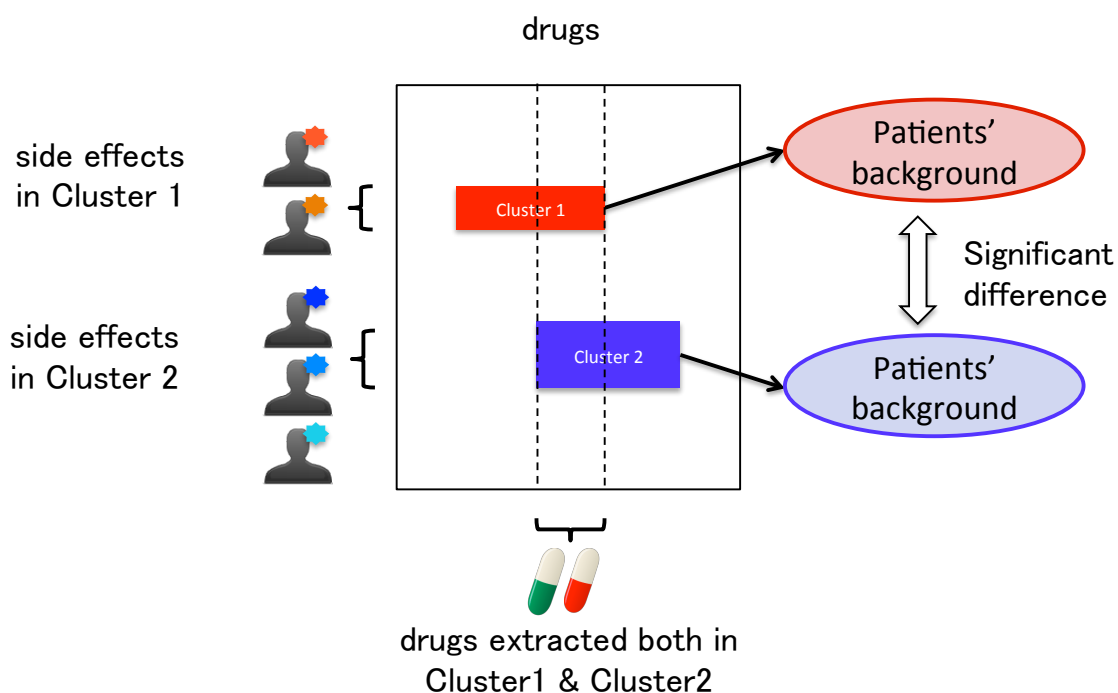


図 20 副作用クラスタ間の比較

この模式図では、Cluster1 と Cluster2 は薬物を 2 つ共通に持つ一方、副作用には共通部分は無い。このようなクラスタ対で患者の背景を比較した。

^{*16} ただし、これらのクラスタ対は共通の薬物を少なくとも一つ持つが、クラスタ内の全ての薬物を共有しているわけではない。また、クラスタ対は薬物ごとに換算しているため、重複を含む。

163 個の副作用クラスタに関連した 182,000 件の有害事象報告に関して、FAERS の DEMO ファイルから、患者の性別、年齢、体重の情報を取得し、17,363 対の各ペアで性別、年齢、体重の分布を比較した。その結果、性別分布では 1,606 対で有意な差 (p 値 < 0.05) が見られた。これらの対は、69 個の薬物と関連していた。年齢分布では 2,514 対で有意な差 (p 値 < 0.05) が見られた。これらの対は、80 個の薬物と関連していた。体重分布では、707 対で有意な差 (p 値 < 0.05) が見られた。これらの対は、49 個の薬物と関連していた。この中から、性別、年齢、体重のそれぞれで代表的なクラスタ対を例示する。ここでは、そのような違いが 2 つ以上の薬物で観察されたクラスタ対を選んで議論する。

3.3.6.1 第 148 クラスタと第 163 クラスタでの性別分布の比較 第 148 クラスタと第 163 クラスタの比較では、患者の性別分布に顕著な違いが見られた (図 21)。第 148 クラスタの患者はすべて女性であったのに対し、第 163 クラスタの患者は女性、男性を含んでいることが観察された。この傾向は、ニキビの治療薬であるイソトレチノイン (Isotretinoin) と抗リウマチ薬であるエタネルセプト (Etanercept) で観察された。イソトレチノインについては、感染症や心血管障害などの副作用が知られている (DailyMed)。エタネルセプトについては、感染症、自己免疫疾患、および心臓障害を発現することが知られている (DailyMed)。

本解析の結果では、第 148 クラスタに *Vaginal and vulval infections and inflammations* (膣および外陰部感染および炎症) が抽出され、患者の集団が全て女性であったこととつじつまが合う。他には *Streptococcal infections* (レンサ球菌感染) や *fungus infections* (真菌感染) などの感染症や *Mass conditions* (腫瘍の状態) および *Autoimmune disorders* (自己免疫疾患) が抽出された (図 21 表左)。このことは、これらの副作用が女性に現れやすい傾向を示唆している。本解析で性特異的に抽出された副作用は、一般的には医薬品のデータベースには記載されていないが、文献で女性でのリスク増加が指摘されている [29]。

第 163 クラスタには、*Neurologic visual problems* (神経性視覚障害)、*Visual disorders* (視覚障害)、*Cardiac signs and symptoms* (心徴候および症状) および *Circulatory collapse and shock* (循環虚脱およびショック) などの副作用が現れた。これらの副作用は、第 148 クラスタの場合とは対照的に、性特異的に発生するとは考えにくい。

3.3.6.2 第 106 クラスタと第 116 クラスタでの年齢分布の比較 第 106 クラスタと第 116 クラスタの比較では、患者の年齢分布に有意差が認められた (図 22)。この傾向は、避妊薬のレボノルゲストレル (Levonorgestrel) と抗腫瘍剤であるインターフェロン -1a

(Interferon beta-1a) に観察された。レボノルゲストレルは、黄体ホルモン作用物質であり、避妊、月経障害の制御、および子宮内膜症の治療のために使用される。インターフェロン- β は、多発性硬化症の寛解や再発を防ぐために使用される。

第 116 クラスタに関連する患者はすべて女性であり、その年齢は、レボノルゲストレルでは 30 歳から 55 歳 (図 22 左)、インターフェロン β -1a では大部分が 30 歳から 50 歳 (図 22 右) で、いずれも中年期の年齢にあった。第 116 クラスタに抽出された副作用は子宮や卵巣などの良性・悪性新生物などであった (図 22 表右)。黄体ホルモン作用物質を含むホルモン療法による卵巣癌および乳癌のリスクは、これらの疾患を既往症として持つ患者 (DailyMed) や、閉経後の女性 [2, 14, 25, 59] で報告されている。本解析の結果は、閉経以前の中年期の女性でも、レボノルゲストレルまたはインターフェロン β -1a のいずれかの使用による子宮や卵巣の良性・悪性新生物のリスクを示唆している。しかしながら、本解析で観察された年齢分布は子宮癌や卵巣癌の好発年齢と重なるため、真に薬物の有害反応であるか否かは定かでない。

第 106 クラスタでは、第 116 クラスタの患者と比較して、レボノルゲストレルでは低い年齢分布 (p 値=2.6E-6) を示し、インターフェロン β -1a では 20 歳代から 70 歳代までのより広範囲な年齢分布 (有意な年齢差の p 値=6.2E-10) を示した。第 106 クラスタでは、*disturbances in consciousness* (意識障害) and *neurological signs and symptoms* (神経学的徴候および症状) などの神経系障害の他、*dyssomnias* (睡眠異常), "*anxiety symptoms* (不安症状), "*emotional and mood disturbances*" (感情および気分障害), "*increased physical activity levels*" (身体の活動性レベルの亢進), "*fluctuating mood symptoms*" (気分変動症状), "*mood alterations with manic symptoms*" (躁症状を伴う気分変動) などの精神疾患に関係のある副作用が抽出された (図 22 表左)。これらの副作用は、レボノルゲストレルでは若い年齢層の患者に多く観察された、神経系障害や精神疾患関係の副作用の若い年齢層での発現傾向が示唆された。インターフェロン β -1a は精神医学的障害を発現することが報告されているが (DailyMed) 我々の結果では年齢に関する可能性は低いことが示唆された。

3.3.6.3 第 108 クラスタと第 156 クラスタでの体重分布の比較 第 108 クラスタおよび第 156 クラスタの比較では、患者の体重分布に有意差が見て取れた (図 23)。この傾向は抗うつ剤であるパロキセチンと、抗炎症剤であるロフェコキシブに観察された。第 108 クラスタの患者が主に成人であったのに対し、第 156 クラスタの患者は主に新生児があったと考えられる (体重がすべて 5 キロ未満であったため) (p 値=1.6E-7)。第 108 クラスタでは *cardiac signs and symptoms* (心徴候および症状)、*circulatory collapse and*

shock (循環虚脱およびショック)などの副作用が抽出された(図 23 表左)。第 156 クラスタでは *non-site specific injuries* (部位不明の損傷)、*inner-ear signs and symptoms* (内耳徴候および症状)、*vertigos* (回転性めまい)、*myocardial disorders* (心筋障害)、*umbilical hernias* (腹壁の状態)などの副作用が抽出された(図 23 表右)。これは、妊婦によるパロキセチンの服用が胎児にこれらの副作用を引き起こすことを示唆している。実際、乳児の心血管系の欠陥のリスクは妊娠中の女性における胎児のパロキセチン曝露の結果として報告されている(DailyMed)。ロフェコキシブは、第 108 クラスタの患者では、第 156 クラスタの患者よりも相対的に低い体重を示した(p 値=3.0E-3)。本結果に関する DailyMed などでの報告は無く、体重の違いが異なる副作用の発現にどう寄与しているかは不明であるが、両クラスタ間で異なる副作用が観察された要因が体重である可能性を示唆している。

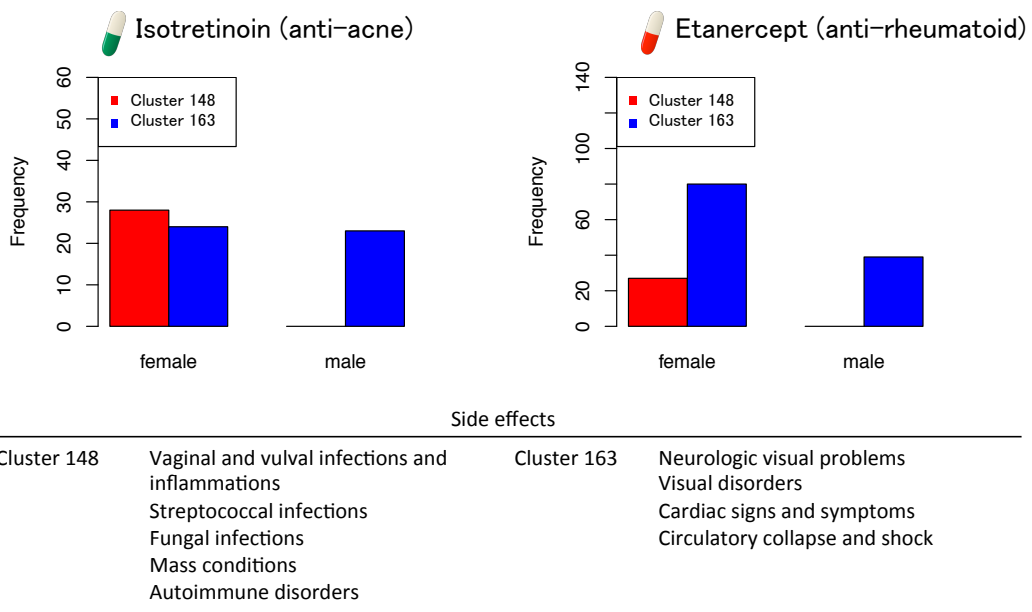


図 21 副作用発現の性別による違い (Mizutani ら [44] より改変)
 第 148 クラスタ (赤) と第 163 クラスタ (青) で患者の性別分布を比較した。第 148 クラスタの患者は全て女性であった。一方、第 163 クラスタは男性、女性を含んでいた。イソトレチノイン (Isotretinoin) とエタネルセプト (Etanercept) で性別分布の差が示された。各クラスタに抽出された副作用を表に記載した。

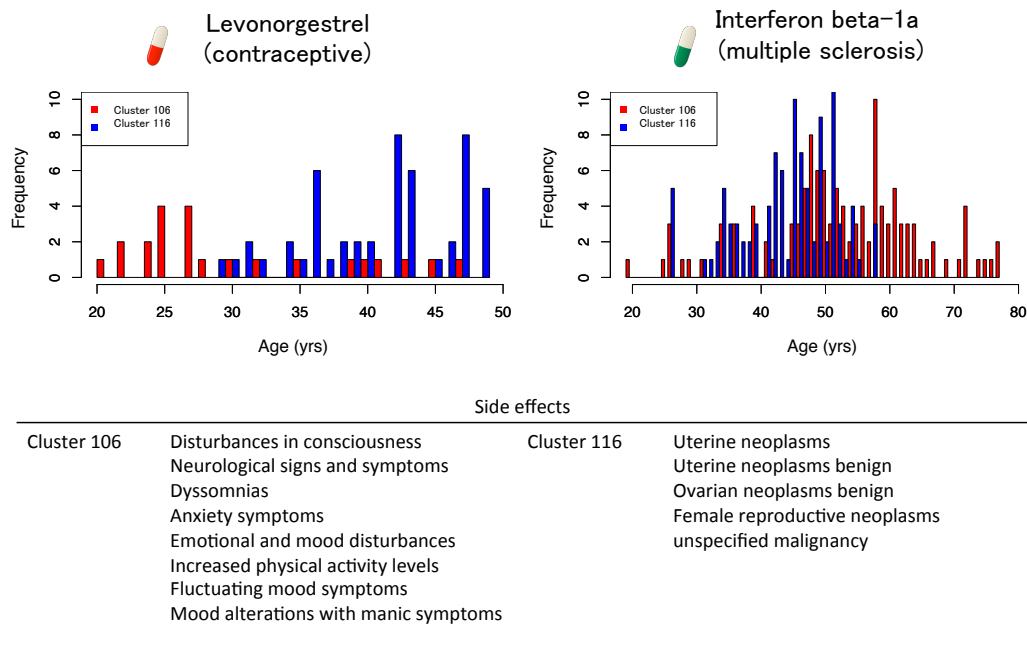
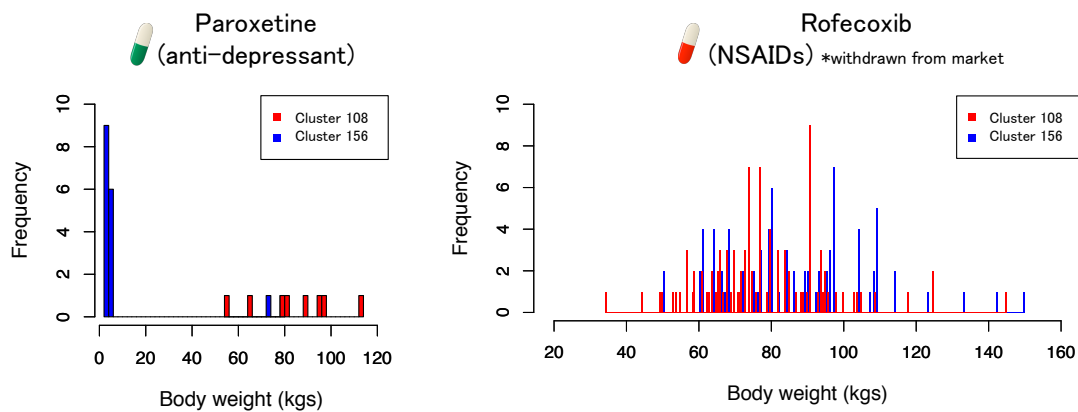


図 22 副作用発現の年齢による違い (Mizutani ら [44] より改変)
 第 106 クラスタ (赤) と第 116 クラスタ (青) で患者の年齢分布を比較した。レボノ
 ルゲストレル (Levonorgestrel) (p 値 = 2.6E-6) インターフェロン β 1-a (Interferon
 beta-1a) (p 値 = 6.2E-10) で年齢分布の有意差が示された。各クラスタに抽出された
 副作用を表に記載した。



Side effects

Cluster 108	Cardiac signs and symptoms Circulatory collapse and shock	Cluster 156	Non-site specific injuries Inner ear signs and symptoms Vertigos Myocardial disorders Umbilical hernias
-------------	--	-------------	---

図 23 副作用発現の体重による違い (Mizutani ら [44] より改変)
 第 108 クラスター (赤) と第 156 クラスター (青) で患者の体重分布を比較した。パロキセチン (Paroxetine)(p 値=1.6E-7) とロフェコキシブ (Rofecoxib)(p 値=3.0E-3) ので体重分布の有意差が示された。各クラスターに抽出された副作用を表に記載した。

3.4 まとめと今後の展望

研究 II では、市販後調査による大規模有害事象報告データを元に、データマイニングの手法であるバイクラスタリングを用いて、発生頻度の高い副作用の集合と、それに関連する薬物の集合を、副作用クラスタとしてを同定した。同定された 163 個の副作用クラスタを評価した結果、145 個の副作用クラスタでは、薬物の適応症以外の共通の機構により、医学的に類似した副作用が発生するケースの説明可能性が示唆された。さらに、副作用クラスタのペアワイズな比較により、同一の薬物による副作用の発生の違いが患者の生理学的性質により説明されるケースを検出した。本研究で使用した手法と、本法を用いて得られた結果について、今後の展望を述べる。

3.4.1 方法に関する展望

バイクラスタリングは、解析対象に関する事前知識に依らず、データ値のみで対象をクラスタリングする方法である。その利点は、従来の仮説駆動型の統計手法と異なり、仮説そのものを発見（マイニング）できることにある。そのため、検出される情報は、全く未知であることもあり、新知見を獲得するのに有効である一方、データ解析には十分に注意を払わなければならない。本研究では、ノイズの多いデータ行列に対しても適用可能であると言われているアルゴリズムを用い、さらにパラメータ調整を注意深く設定してクラスタを同定した。その結果、データから薬物や副作用に存在する既存の知識や多様性をそれほど失うことなく、有用な情報を抽出することができた。

3.1.1 でも述べたように、薬剤疫学の分野では、自発報告システムに登録された市販後大規模報告から未知の副作用を検出することを目的として、様々なシグナル検出法が開発されている。中でも Dumouchel により開発された gamma Poisson shrinker (GPS)[16] は、それ以前から使われていた proportional reporting ratio (PRR) を改良したもので、薬剤疫学の研究分野では定評がある。シグナル検出法では、薬物と副作用のペアを出力とし、バイクラスタリングとは出力形式が異なる。しかしながら、2 手法の利点を比較するには、出力結果の質的、量的な評価を行うことが必要であり、今後の課題となる。

自発報告システムには、その性質上、2 種類のバイアスが伴うと言われている。一つ目のバイアスは、サンプリングによる分散 (sampling variance) と呼ばれ、良く知られた被疑薬が過剰に報告されることによって生じるバイアスのことを指す。二つ目のバイアスは、観測選択効果 (selection bias) と呼ばれ、サンプル集団に存在する層化 (stratification) (患者の性別、年齢、病歴、国籍などの他、報告者の職業、報告団体の種類など) がある

ために生じるバイアスのことを指す。このバイアスのため、薬物を投与された集団とコントロールの間の正当な比較が困難であるといわれている。この2種類のバイアスは、従来のシグナル検出法によっても完全には排除することは不可能であるといわれている。Tatonettiらはこの2種類の注意点を指摘し、それらを排除する新しい手法を提案している [62]。バイクラスタリングでは通常このような問題を回避する枠組みはないが、これらの統計的な問題を回避することは今後の課題となるであろう。

3.4.2 結果に関する展望

本研究で同定された163個の副作用クラスタのうちの145個(89%)では、副作用の医学的類似性が確認され、薬物の適応症以外に共通の機構が存在することが示唆された(3.3.4)。共通の分子機構として、3つの副作用クラスタについてその機序を詳細に検証した(表12、図19)。

第17クラスタの「出血」には、3つのタイプの薬物による抗血液凝固反応が原因であることが推定された。この副作用は主に消化管と神経系で起きていたが、その理由は明らかではない。このことを明らかにするには、薬物の用法に関する情報を用いる必要がある。FAERSには、薬物の投与ルートに登録する項目があるため、この情報を用いることにより類似した副作用発現の臓器別の発症機序を解析する手がかりとなりうる。

第29クラスタでは、シルデナフィルが眼疾患を引き起こしている原因として、その標的タンパク質であるcGMPホスホジエステラーゼが網膜を含む多様な組織で発現していることが原因であると推定できた。実際、シルデナフィルのホスホジエステラーゼ阻害による視覚障害がマウスで確認され、その障害の程度がphosphodiesterase 6 (PDE6)変異マウスではより顕著であったとの報告がある[48]。すなわち、報告された副作用の発生が、本来の治療臓器以外での薬理作用によるものであることが確認された。しかしながら、このような検証を全ての薬物や標的タンパク質で行うことは難しい。タンパク質の発現には組織得意的な発現機構があり、近年、マイクロアレイや次世代シーケンサにより組織別に遺伝子発現を調査したデータが取られているため、標的タンパク質の組織の発現解析を行うことで、この推定をより確かなものにできると考えられる。

本解析では、患者の生理学的性質の違いにより発現の異なる副作用のケースを検出し、その機序を推定した。本研究では、患者の個人差を性別・年齢・体重で特徴付け、同じ薬物でも女性のみが発現が見られる副作用や、妊娠中の服用により胎児に重篤な副作用が起きたと思われるケースが検出された。しかしながら、個人差による副作用の発現機構を分子レベルで解明するためには、生理学的な情報だけでは限界がある。なぜなら、薬物の副作用には遺伝的な原因によるものが少なくないからである。良く知られている例として、

クロピドグレルが挙げられる。クロピドグレルは、代謝酵素である CYP2C19 の多型による有害反応が知られている。PharmGKB は、医薬品に関する情報を網羅的に収集したデータベースであるが、特に、薬物代謝酵素や標的タンパク質の遺伝的多型に関する知見を収集し提供している。将来的には、遺伝的多型のデータを市販後調査のデータに統合して解析することが理想であるが、現状では著者の知る限りそのような公共データベースはまだ無い。

このように、本研究で用いたデータソースや手法には様々な意味での発展が考えられる。市販後有害事象データの解析に生物学的な知見を統合することにより、薬物疫学の分野に新たな解析法を提供することが期待できる。

第4章 全体のまとめ

医薬品としての薬物は、病気の症状を治療・緩和する効果（主作用）を持つが、一方で意図せぬ弊害（副作用）を引き起こすこともある。副作用の発現は患者の生活の質を低下させるだけでなく、治療の中断につながることもあり、医薬品の安全性や有効性を考える上で回避しなければならない問題である。副作用の発現の原因になるような毒性や有害反応は、前臨床試験および臨床試験で検証されるにも関わらず、副作用発現への全容解明には未だほど遠い。その主な理由として、1つ目は前臨床試験での *in vitro* 実験および動物実験とヒトでの有害反応の出方に隔たりがあること、2つ目は臨床試験（治験）での個人差を考慮した副作用発現の検出に限界があることが考えられる。また、副作用発現の機序については、一部の市場撤退薬物や重篤な副作用を引き起こす可能性のある薬物を除いては、その全容解明には未だ至っていない。薬物の副作用を創薬段階で予測し回避することは重要課題な課題であるが、それと同時に、市販後に報告された副作用例から統計的手法を用いて重要な情報を抽出し、その発現機構を解明することも重要な研究課題である。近年の医薬品関連データベースの発達に伴い、創薬段階および市販後に調査収集された大規模データから分子作用機序を推定する計算機科学的手法が提案されてきた。しかしながら今までの研究では、薬物の化学情報や標的分子への直接的な薬理作用に注目されて来たため、生体の複雑な階層構造を考慮した機序の解明には十分に至っていなかった。また、個人差を加味した副作用の種類の違いや発現の度合いは、あまり考慮されて来なかった。

本研究では、この2つの問題解明を目的とし、それぞれに適切な手法を適用して、副作用発現機序の解明および予測の為に解析を行った。研究 I では、標的タンパク質の相互作用情報を用いることで分子経路を介した副作用発現をより深く理解することができるのではないかと考え、薬物と標的タンパク質の相互作用情報と副作用を関連付ける方法を提案した。その結果、特定の副作用と顕著な関係にある標的タンパク質同士は互いに共通の分子ネットワークを形成している傾向が強いことが示された。さらに提案手法を未知の副作用予測に応用したところ、化学構造情報を用いた既存の手法より高い予測精度を達成することができた。研究 II では、類似した発生頻度を示す副作用は互いに共通の発現機序を持つのではないかと考え、大規模な市販後報告データを用いて副作用クラスタを同定した。その結果、異なる薬物でも、共通の薬理作用により、類似した副作用を起こす例を見いだした。さらに、同じ薬物でも、患者の背景により発症する副作用の異なる例を検出することができた。

生体は、複雑な階層構造によって精密に制御されているため、薬理作用や副作用のパ

ターンも様々である。薬理作用の影響は、分子、細胞、組織、臓器、個体など様々なレベルで観測される。そのため、医薬品関連データは今後さらに多様性を持ち、かつデータ量も多くなると考えられる。本研究で提案した解析法は、このようなデータソースから重要な相関関係を検出するための有効な手段であると考えられる。また、本解析法は、個別化医療の発展に伴い、将来的には、個人間の違いを考慮した薬剤疫学データの解析への応用にも期待できる。

謝辞

本研究は、京都大学化学研究所バイオインフォマティクスセンター化学生命科学領域（緒方研究室）において、五斗進准教授の指導の元に行われました。五斗准教授には、研究を進める上で重要な指針を与えて頂き、本稿および主論文の基礎となる2本の論文の執筆にあたり丁寧に御指導を頂きました。深く感謝致します。

同研究室の緒方博之教授には、本稿執筆にあたり指針を頂きました。深く感謝致します。

同センターの金久實特任教授には、バイオインフォマティクスへのビジョンを与えて頂くと共に、本研究を進める上で大変貴重な助言を頂きました。有り難うございました。

第2章における標的タンパク質と副作用の関連解析は、九州大学生体防御医学研究所システムコホート学分野の山西芳裕准教授、パリ国立高等鉱業学校（Mines ParisTech）のVéronique Stoven教授とEdouard Pauwels氏（現・Israel Institute of Thechnology 博士研究員）との共同研究として行われました。山西准教授には研究の開始から熱意ある御指導を頂き、方法論から論文の執筆に至るまで幅広く御指導頂きました。深く感謝致します。

この共同研究はパリのMines ParisTechでのJean-Philippe Vert教授の研究室滞在中に開始されましたが、滞在を支援して下さったInternational Training Program（ITP）のプロジェクトリーダーである馬見塚拓教授に感謝申し上げます。

小寺正明助教（現・東京工業大学生命理工学部大学院生命理工学研究科講師）には、同研究室に所属した当初からプログラミングの基礎を始め、研究の進め方から論文の執筆に至るまで細かく御指導頂きました。深く感謝致します。時松敏明助教、守屋勇樹氏（現・DBCLS ライフサイエンス統合データベースセンター研究員）、武藤愛氏（現・奈良先端科学技術大学院大学システム微生物学助教）、Gemma M. Kirwan氏、Nicolas Joannin氏には研究内容および技術的な面で様々な助言を頂きました。

同研究室の研究員である清水祐吾氏、山本留美子氏、学生の西村陽介氏、三原知子氏には、日頃から研究生活において協力を頂きました。特に卒業生である野呂洋介氏には、第3章の研究の基礎となるデータの作成と解析手法に関して助力を頂きました。また、バイオインフォマティクスセンターの先生方、湯川かえで秘書、KEGGの方々、スーパーコンピュータラボラトリーの方々にもお世話になりました。この場を借りて感謝申し上げます。

最後に、本研究を進めるにあたり日頃から支えて下さった友人、家族に感謝の意を表したいと思います。

参考文献

- [1] W. Ageno, A. Callus, A. Wittkowsky, M. Crowther, E. Hylek, and G. Palareti. Oral anticoagulant therapy: Antithrombotic therapy and prevention of thrombosis, 9th ed: American college of chest physicians evidence-based clinical practice guidelines. *CHEST Journal*, 141(2):e44S–e88S, 2012.
- [2] G. L. Anderson, M. Limacher, A. R. Assaf, T. Bassford, S. A. A. Beresford, H. Black, D. Bonds, R. Brunner, R. Brzyski, B. Caan, R. Chlebowski, D. Curb, M. Gass, J. Hays, G. Heiss, S. Hendrix, B. V. Howard, J. Hsia, A. Hubbell, R. Jackson, K. C. Johnson, H. Judd, J. M. Kotchen, L. Kuller, A. Z. LaCroix, D. Lane, R. D. Langer, N. Lasser, C. E. Lewis, J. Manson, K. Margolis, J. Ockene, M. J. O’Sullivan, L. Phillips, R. L. Prentice, C. Ritenbaugh, J. Robbins, J. E. Rossouw, G. Sarto, M. L. Stefanick, L. Van Horn, J. Wactawski-Wende, R. Wallace, S. Wassertheil-Smoller, and Women’s Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women’s Health Initiative randomized controlled trial. *JAMA : the journal of the American Medical Association*, 291(14):1701–1712, 2004.
- [3] J. K. Aronson, editor. *Side Effects of Drugs Annual*, volume 29. Elsevier, 2007.
- [4] N. Atias and R. Sharan. An algorithmic framework for predicting side-effects of drugs. *Journal of Computational Biology*, 18(3):207–218, 2011.
- [5] J. Backman. Gemfibrozil greatly increases plasma concentrations of cerivastatin. *Clinical Pharmacology and Therapeutics*, 72(6):685–691, 2002.
- [6] J. P. F. Bai and D. R. Abernethy. Systems Pharmacology to Predict Drug Toxicity: Integration Across Levels of Biological Organization*. *Annual Review of Pharmacology and Toxicology*, 53(1):451–473, 2013.
- [7] A. Bate and S. J. W. Evans. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiology and Drug Safety*, 18(6):427–436, 2009.
- [8] A. Bender, J. Scheiber, M. Glick, J. W. Davies, K. Azzaoui, J. Hamon, L. Urban, S. Whitebread, and J. L. Jenkins. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem*, 2(6):861–873, 2007.
- [9] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practi-

- cal and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, 50:289–300, 1995.
- [10] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, and J. P. Overington. The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, 42(Database issue):D1083–90, 2014.
- [11] S. Bergmann, J. Ihmels, and N. Barkai. Iterative signature algorithm for the analysis of large-scale gene expression data. *Physical Review E*, 67(3):031902, 2003.
- [12] M. Campillos, M. Kuhn, A. Gavin, L. Jensen, and P. Bork. Drug target identification using side-effect similarity. *Science*, 321(5886):263–266, 2008.
- [13] F. K. L. Chan and D. Y. Graham. Prevention of non-steroidal anti-inflammatory drug gastrointestinal complications - review and recommendations based on risk assessment. *Alimentary Pharmacology and Therapeutics*, 19(10):1051–1061, 2004.
- [14] R. T. Chlebowski, S. L. Hendrix, R. D. Langer, M. L. Stefanick, M. Gass, D. Lane, R. J. Rodabough, M. A. Gilligan, M. G. Cyr, C. A. Thomson, J. Khandekar, H. Petrovitch, A. McTiernan, and WHI Investigators. Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women’s Health Initiative randomized trial. *JAMA : the journal of the American Medical Association*, 289(24):3243–3253, 2003.
- [15] C. Dessy, I. Kim, C. L. Sougnez, R. Laporte, and K. G. Morgan. A role for map kinase in differentiated smooth muscle contraction evoked by alpha-adrenoceptor stimulation. *American Journal of Physiology - Cell Physiology*, 275(4):C1081–C1086, 1998.
- [16] W. Dumouchel. Bayesian data mining in large frequency tables, with an application to the fda spontaneous reporting system. *The American Statistician*, 53(3):177–190, 1999.
- [17] S. Ekins and J. A. Erickson. A pharmacophore for human pregnane X receptor ligands. *Drug metabolism and disposition: the biological fate of chemicals*, 30(1):96–99, Jan. 2002.
- [18] K. Finlayson, H. Witchelb, J. McCullocha, and J. Sharkey. Acquired QT interval prolongation and hERG: implications for drug discovery and development. *European Journal of Pharmacology*, 500:129–142, 2004.

- [19] A. Fliri, W. Loging, P. Thadeio, and R. Volkmann. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nature Chemical Biology*, 1:389–397, 2005.
- [20] A. Fliri, W. Loging, P. Thadeio, and R. Volkmann. Analysis of system structure-function relationships. *ChemMedChem*, 2:1774–1782, 2007.
- [21] W. Fujibuchi, S. Goto, H. Migimatsu, I. Uchiyama, A. Ogiwara, Y. Akiyama, and M. Kanehisa. DBGET/LinkDB: an integrated database retrieval system. *Proceedings on Pacific Symposium on Biocomputing*, pages 683–694, 1998.
- [22] K. M. Giacomini, R. M. Krauss, D. M. Roden, M. Eichelbaum, M. R. Hayden, and Y. Nakamura. When good drugs go bad. *Nature*, 446(7139):975–977, 2007.
- [23] S. Gimenez-Roldan and D. Mateo. Cinnarizine-induced parkinsonism. susceptibility related to aging and essential tremor. *Clinical Neuropharmacology*, 14(2):156–164, 1991.
- [24] C. R. Gordon, A. Gonen, Z. Nachum, I. Doweck, O. Spitzer, and A. Shupak. The effects of dimenhydrinate, cinnarizine and transdermal scopolamine on performance. *Journal of Psychopharmacology*, 15(3):167–172, 2001.
- [25] D. Grady, T. Gebretsadik, K. Kerlikowske, V. Ernster, and D. Petitti. Hormone replacement therapy and endometrial cancer risk: a meta-analysis. *Obstetrics and Gynecology*, 85(2):304–313, 1995.
- [26] S. Günther, M. Kuhn, M. Dunkel, M. Campillos, C. Senger, E. Petsalaki, J. Ahmed, G. Urdiales, A. Gewiess, L. Jensen, R. Schneider, R. Skoblo, R. Russell, P. Bourne, P. Bork, and R. Preissner. Supertarget and matador: resources for exploring drug-target relationships. *Nucleic Acids Research*, 36:D919–D922, 2008.
- [27] R. Harpaz, S. Vilar, W. Dumouchel, H. Salmasian, K. Haerian, N. H. Shah, H. S. Chase, and C. Friedman. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *Journal of the American Medical Informatics Association*, 20(3):413–419, 2013.
- [28] N. Hecker, J. Ahmed, J. von Eichborn, M. Dunkel, K. Macha, A. Eckert, M. K. Gilson, P. E. Bourne, and R. Preissner. SuperTarget goes quantitative: update on drug-target interactions. *Nucleic Acids Research*, 40(D1):D1113–D1117, 2011.
- [29] M. V. Hernández, R. Sanmartí, J. D. Cañete, M. A. Descalzo, M. Alsina, L. Carmona, J. J. Gomez-Reino, and on behalf of the BIOBADASER 2.0 Study Group.

- Cutaneous adverse events during treatment of chronic inflammatory rheumatic conditions with tumour necrosis factor antagonists: Study using the spanish registry of adverse events of biological therapies in rheumatic diseases. *Arthritis Care & research*, 65(12):2024–2031, 2013.
- [30] J. Horowitz, Y. Almog, A. Wolf, G. Buckan, and O. Geyer. Ophthalmic complications of dental anesthesia : Three new cases. *Journal of Neuroophthalmology*, 25(2):95–100, 2005.
- [31] J. Ihmels, S. Bergmann, and N. Barkai. Defining transcription modules using large-scale gene expression data. *Bioinformatics*, 20(13):1993–2003, 2004.
- [32] M. Iskar, G. Zeller, P. Blattmann, M. Campillos, M. Kuhn, K. H. Kaminska, H. Runz, A.-C. Gavin, R. Pepperkok, V. van Noort, and P. Bork. Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. *Molecular Systems Biology*, 9(662):1–13, 2013.
- [33] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, 42(Database issue):D199–D205, 2014.
- [34] M. Keiser, B. Roth, B. Armbruster, P. Ernsberger, J. Irwin, and B. Shoichet. Relating protein pharmacology by ligand chemistry. *Nature Biotechnology*, 25:197–206, 2007.
- [35] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, and B. K. Shoichet. Relating protein pharmacology by ligand chemistry. *Nature Biotechnology*, 25(2):197–206, 2007.
- [36] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, and D. S. Wishart. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Research*, 39(Database issue):D1035–1041, 2011.
- [37] M. Kuhn, M. Al Banhaabouchi, M. Campillos, L. J. Jensen, C. Gross, A.-C. Gavin, and P. Bork. Systematic identification of proteins that elicit drug side effects. *Molecular Systems Biology*, 9:1–9, 2013.
- [38] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, and P. Bork. A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology*, 6:343, 2010.
- [39] M. Kuhn, D. Szklarczyk, A. Franceschini, C. von Mering, L. J. Jensen, and

- P. Bork. STITCH 3: zooming in on protein-chemical interactions. *Nucleic Acids Research*, 40(Database issue):D876–80, Jan. 2012.
- [40] J. Lam, S. Schulman, D. M. Witt, P. O. Vandvik, F. Qayyum, and A. M. Holbrook. Anticoagulation control with daily low-dose vitamin k to reduce clinically adverse outcomes and international normalized ratio variability: A systematic review and meta-analysis. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 33(11):1184–1190, 2013.
- [41] Q. Li, T. Cheng, Y. Wang, and S. Bryant. Pubchem as a public resource for drug discovery. *Drug Discovery Today*, 23-24(9):1052–1057, 2010.
- [42] E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, B. K. Shoichet, and L. Urban. Large-scale prediction and testing of drug activity on side-effect targets. *Nature*, 486(7403):361–367, 2012.
- [43] M. Mathews, S. Gratz, B. Adetunji, V. George, M. Mathews, and B. Basil. Antipsychotic-induced movement disorders: Evaluation and treatment. *Psychiatry*, 2(3):36–41, 2005.
- [44] S. Mizutani, Y. Nono, M. Kotera, and S. Goto. Pharmacoepidemiological characterization of drug-induced adverse reaction clusters towards understanding of their mechanisms. *Computational Biology and Chemistry*, 50:50–59, 2014.
- [45] S. Mizutani, E. Pauwels, V. Stoven, S. Goto, and Y. Yamanishi. Relating drug–protein interaction network with drug side effects. *Bioinformatics*, 28(18):i522–i528, 2012.
- [46] J. A. Morón, A. Brockington, R. A. Wise, B. A. Rocha, and B. T. Hope. Dopamine uptake through the norepinephrine transporter in brain regions with low levels of the dopamine transporter: evidence from knock-out mouse lines. *The Journal of Neuroscience*, 22(2):389–395, 2002.
- [47] G. M. Nidhi, J. W. Davies, and J. L. Jenkins. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *Journal of Chemical Information and Modeling*, 46(3):1124–1133, 2006.
- [48] L. Nivison-Smith, Y. Zhu, A. Whatham, B. V. Bui, E. L. Fletcher, M. L. Acosta, and M. Kalloniatis. Sildenafil alters retinal function in mouse carriers of Retinitis Pigmentosa. *Experimental Eye Research*, 128:43–56, 2014.

- [49] J. C. O’Keefe and J. E. Creamer. Efficacy of nisoldipine combined with beta-adrenergic-blocking drugs in the treatment of chronic stable angina. *Clinical Cardiology*, 10(6):345–350, 1987.
- [50] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins. How many drug targets are there? *Nature reviews. Drug discovery*, 5(12):993–996, 2006.
- [51] J. Papay, N. Yuen, G. Powell, M. Mockenhaupt, and T. Bogenrieder. Spontaneous adverse event reports of Stevens-Johnson syndrome/toxic epidermal necrolysis: detecting associations with medications. *Pharmacoepidemiology and Drug Safety*, 21(3):289–296, 2012.
- [52] E. Pauwels, V. Stoven, and Y. Yamanishi. Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC Bioinformatics*, 12(1):169, 2011.
- [53] O.A. C. Petroff, F. Hyder, D. L. Rothman, and R. H. Mattson. Effects of gabapentin on brain GABA, homocarnosine, and pyrrolidinone in epilepsy patients. *Epilepsia*, 41(6):675–680, 2000.
- [54] T. Prueksaritanont. Mechanistic studies on metabolic interactions between gemfibrozil and statins. *Journal of Pharmacology and Experimental Therapeutics*, 301(3):1042–1051, 2002.
- [55] C. Qin, C. Zhang, F. Zhu, F. Xu, S. Y. Chen, P. Zhang, Y. H. Li, S. Y. Yang, Y. Q. Wei, L. Tao, and Y. Z. Chen. Therapeutic target database update 2014: a resource for targeted therapeutics. *Nucleic Acids Research*, 42(Database issue):D1118–D1123, 2014.
- [56] J. Scheiber, B. Chen, M. Milik, S. Sukuru, A. Bender, D. Mikhailov, S. Whitebread, J. Hamon, K. Azzaoui, L. Urban, M. Glick, J. Davies, and J. Jenkins. Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. *Journal of Chemical Information and Modeling*, 49(2):308–317, 2009.
- [57] J. Scheiber, J. Jenkins, S. Sukuru, A. Bender, D. Mikhailov, M. Milik, K. Azzaoui, S. Whitebread, J. Hamon, L. Urban, M. Glick, and J. Davies. Mapping adverse drug reactions in chemical space. *Journal of Medicinal Chemistry*, 52(9):3103–3107, 2009.
- [58] Y. Shitara, M. Hirano, H. Sato, and Y. Sugiyama. Gemfibrozil and its glucuronide inhibit the organic anion transporting polypeptide 2 (OATP2/OATP1B1:SLC21A6)-mediated hepatic uptake and CYP2C8-mediated

- metabolism of cerivastatin: analysis of the mechanism of the clinically relevant drug-drug interaction between cerivastatin and gemfibrozil. *The Journal of Pharmacology and Experimental Therapeutics*, 311(1):228–236, 2004.
- [59] M. L. Stefanick, G. L. Anderson, K. L. Margolis, S. L. Hendrix, R. J. Rodabough, E. D. Paskett, D. S. Lane, F. A. Hubbell, A. R. Assaf, G. E. Sarto, R. S. Schenken, S. Yasmeen, L. Lessin, R. T. Chlebowski, and WHI Investigators. Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *JAMA : the journal of the American Medical Association*, 295(14):1647–1657, 2006.
- [60] T. Tamura, T. Sakaeda, K. Kadoyama, and Y. Okuno. Aspirin- and clopidogrel-associated bleeding complications: Data mining of the public version of the fda adverse event reporting system, aers. *International Journal of Medical Sciences*, 9(6):441–446, 2012.
- [61] N. Tatonetti, T. Liu, and R. Altman. Predicting drug side-effects by chemical systems biology. *Genome Biology*, 10:238–241, 2009.
- [62] N. Tatonetti, P. Ye, R. Daneshjou, and R. Altman. Data-Driven Prediction of Drug Effects and Interactions. *Science Translational Medicine*, 4(125):125–31, 2012.
- [63] M. I. van der Reis, E. C. La Heij, Y. De Jong-Hesse, P. J. Ringens, F. Hendrikse, and J. S. A. G. Schouten. A systematic review of the adverse events of intravitreal anti-vascular endothelial growth factor injections. *Retina, The Journal of Retinal and Vitreous Diseases*, 31(8):1449–1469, 2011.
- [64] S. Vilar, R. Harpaz, H. S. Chase, S. Costanzi, R. Rabadan, and C. Friedman. Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. *Journal of the American Medical Informatics Association*, 18 Suppl 1:i73–i80, 2011.
- [65] J. Walker, H. Houston, S. Miller, and G. Roan. Acute methemoglobinemia secondary to topical benzocaine spray. *Advanced Studies in Medicine*, 3(1):45–48, 2003.
- [66] I. Wallach, N. Jaitly, and R. Lilien. A structure-based approach for mapping adverse drug reactions to the perturbation of underlying biological pathways. *PLoS One*, 5:e12063, 2010.
- [67] M. Whirl-Carrillo, E. McDonagh, J. Hebert, L. Gong, K. Sangkuh, C. Thorn,

- R. Altman, and T. Klein. Pharmacogenomics Knowledge for Personalized Medicine. *Clinical Pharmacology and Therapeutics*, 92(4):414–417, 2012.
- [68] S. Whitebread, J. Hamon, D. Bojanic, and L. Urban. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug Discovery Today*, 10(21):1421–1433, 2005.
- [69] D. A. Williams, W. O. Foye, and T. L. Lemke. *Foye’s Principles of Medicinal Chemistry*. Lippincott, Williams and Wilkins, fifth edition edition, 2002.
- [70] D. Witten, R. Tibshirani, and T. Hastie. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10:515–534, 2009.
- [71] L. Xie, J. Li, L. Xie, and P. Bourne. Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. *PLoS Computational Biology*, 5:e1000387, 2009.
- [72] L. Xie, J. Wang, and P. E. Bourne. In silico elucidation of the molecular mechanism defining the adverse effect of selective estrogen receptor modulators. *PLoS Computational Biology*, 3(11):e217, Nov. 2007.
- [73] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 24:i232–i240, 2008.
- [74] Y. Yamanishi, M. Kotera, M. Kanehisa, and S. Goto. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics*, 26:i246–i254, 2010.
- [75] Y. Yamanishi, J. P. Vert, A. Nakaya, and M. Kanehisa. Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis. *Bioinformatics*, 19 Suppl 1:i323–30, 2003.
- [76] Y. Yu, E. Ricciotti, R. Scalia, S. Y. Tang, G. Grant, Z. Yu, G. Landesberg, I. Crichton, W. Wu, E. Puré, C. D. Funk, and G. A. FitzGerald. Vascular COX-2 modulates blood pressure and thrombosis in mice. *Science Translational Medicine*, 4(132):132–54, 2012.